

Der Einsatz von Statistik-Programmpaketen: Klassifikation und vergleichender Überblick

Sensch, Jürgen

Veröffentlichungsversion / Published Version
Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:
GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:

Sensch, J. (1995). Der Einsatz von Statistik-Programmpaketen: Klassifikation und vergleichender Überblick. *Historical Social Research*, 20(3), 3-62. <https://doi.org/10.12759/hsr.20.1995.3.3-62>

Nutzungsbedingungen:

Dieser Text wird unter einer CC BY Lizenz (Namensnennung) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier:
<https://creativecommons.org/licenses/by/4.0/deed.de>

Terms of use:

This document is made available under a CC BY Licence (Attribution). For more Information see:
<https://creativecommons.org/licenses/by/4.0>

Der Einsatz von Statistik-Programmpaketen: Klassifikation und vergleichender Überblick

*Jürgen Sensch**

Abstract: There is no doubt that computer statistical packages provide scientists with a relatively easy way of dealing with complex data analysis, and that the use of computer-based data analysis has increased dramatically in recent years. Computers play nowadays an essential role in the analysis of data in many disciplines, including Historical Social Research. This text begins with a short history of the development of computer statistical packages. A discussion follows which is devoted to developing an understanding of programmed statistical packages in general. With this background a classification scheme for universal and more specific statistical packages follows in part 1. In part 2 we discuss the advantages and disadvantages of statistical packages. At present many packages are dealing with simple and complex data analyses. The earliest packages were designed to be used with large main-frame computers. Nowadays, however, many of them are available on small machines known as personal computers (PCs). In part 3 we therefore give an overview of selected packages with a wide range of statistical procedures and graphical tools suitable for PCs. In part 4 we compare these packages following selected criterias (system characteristics, data management, statistical procedures, graphic tools). The last part deals with problems of artificial intelligence within the statistical computing area. We discuss these purposes more profoundly, and give some of the historical backgrounds. Further objectives are to discuss some knowledge based systems and to show there limitations until today.

* Address all communications to Jürgen Sensch, Center for Historical Social Research, Liliencronstr. 6, D-50931 Köln, Germany; e-mail: sensch@ibm.za.uni-koeln.de.

Was die »traditionelle« Anwendung von Statistik von der »modernen« innerhalb der Historischen Sozialforschung unterscheidet, ist zweifellos der routinemäßige Einsatz der elektronischen Datenverarbeitung bei der Datenaufbereitung und der Datenanalyse. Sie hat die Verbreitungs- und Erkenntnismöglichkeiten der Quantifizierung außerordentlich erweitert.

Wann der Einsatz von EDV notwendig und sinnvoll ist, kann hier nicht im Detail geklärt werden. Allein die zeitlich immer kürzeren Entwicklungsschübe von Hard- und Software auf dem Personalcomputer-Markt, verbunden mit einer ständig wachsenden individualisierten Verbreitung der EDV über die bisherigen Großrechenzentren als ehemals EDV-Monopolisten hinaus, machen zumindest jede Kosten-Nutzen-Analyse fragwürdig.¹ Löst man sich von dem Kostenaspekt, dann läßt sich folgende Faustregel formulieren:

Liegen historische Quellen in quantifizierbarer Form vor und sind sie mit herkömmlichen Methoden nicht mehr überschaubar oder bearbeitbar, dann ist der Einsatz der EDV auf jeden Fall angebracht. Gerade die EDV bietet die Möglichkeit, auch für große Datenmengen problemadäquate Datenaufbereitungs- sowie statistische Auswertungstechniken einzusetzen.² Eine »manuelle« Aufbereitung und Auswertung läßt dagegen nur sehr einfache Auswertungsstrategien zu, die nicht nur sehr begrenzte Analyseverfahren beinhalten, sondern auch unzutreffende Ergebnisse hervorbringen können.

Mit dem Einsatz der EDV beginnt die Domäne der statistischen Auswertungssysteme, die dem Fachwissenschaftler als Werkzeug der Datenanalyse dienen. Sie stellen dem Anwender eine Sammlung von aufeinander abgestimmten Programmen unter einer gemeinsamen Benutzeroberfläche zur Verfügung. Diese Programme selbst sind in einer höheren Programmiersprache wie z.B. FORTRAN (»Formular Translation«) verfaßt und bilden gemeinsam eine Art Metasprache. Insbesondere sind rechenintensive Auswertungsmethoden, das Auswerten großer Datenbestände und das mehrfache Auswerten eines Datenbestandes ohne statistische Auswertungssysteme in der Regel nicht praktikabel. Statistische Auswertungssysteme werden für die beschreibende und schließende Statistik sowie für die explorative Datenanalyse eingesetzt. Die Unterstützung umfaßt sämtliche Arbeitsschritte statistischer Untersuchungen:

- Datenaufnahme und - Verwaltung,

¹ In den vergangenen Jahren ist die PC-Technologie nicht nur leistungsfähiger geworden (im Hinblick auf die Datenverarbeitungsgeschwindigkeit und Speicherkapazität), sondern die Marktentwicklung hat auch zu einem starken Preisverfall im Hardwarebereich geführt.

² Darüberhinaus spricht noch ein wesentliches Argument für die Erstellung maschinenlesbarer Daten: in dieser Form stehen die Daten einer Sekundäranalyse offen, lassen sich leicht für eine vergleichende Studie heranziehen oder mit anderen Teildatensätzen gleicher Grundgesamtheiten maschinell zu einem neuen Gesamtdatensatz verketten.

- Auswertung (Datenanalyse) und
- Ergebnispräsentation.

In der Phase der Datenaufnahme können Datenerfassung, Datenaufbereitung und -speicherung, die Überprüfung oder das Monitoring mit Hilfe statistischer Auswertungssysteme erfolgen. In der Auswertungsphase können die Daten beschrieben und mit Hilfe statistischer Verfahren analysiert werden. Die Ergebnispräsentation umfaßt Graphiken und tabellarische Darstellungsformen.

Ein statistisches Auswertungssystem besteht aus einer Datenbank, welche die auszuwertenden Daten enthält und aus einer Programmbank, die Programme zur statistischen Auswertung bereitstellt. Die Daten- und Programmbank werden verwaltet durch ein Daten- und Programmverwaltungssystem. Für den Anwender sind die Schnittstellen zum statistischen Auswertungssystem wichtig. Eine dieser Schnittstellen ist die sogenannte Kommandosprache, über die der Anwender seine Datenbestände beschreibt (z.B. Vergabe von Variablennamen, Vergabe von Etiketten für die einzelnen Variablenausprägungen etc.) und die Programme zur Auswertung der Datenbestände aktiviert (durch den Aufruf von speziellen Auswertungsprozeduren). Die Entwicklung der statistischen Auswertungssysteme begann in den 60er Jahren. Es lassen sich grob vier Phasen unterscheiden:

- Zunächst wurden EDV-Programme für spezielle Auswertungsmethoden entwickelt. Programmierer entwickelten für die Lösung eines statistischen Problems einen Lösungsalgorithmus, der in einem zweiten Schritt in ein Programm umgesetzt wurde. Programme sind dabei eine Menge von Anweisungen (oder Arbeits Vorschriften), mit deren Hilfe die erhobenen Daten im Computer verarbeitet werden; Programmiersprachen sind eigens für die Abfassung solcher Programme entwickelte Sprachen (z.B. in FORTRAN IV, PL1, C, ALGOL etc.). Mit Hilfe dieser Programmiersprachen können nun Programme geschrieben werden, die dem Computer anweisen, einen Satz von Daten in ganz bestimmter Weise zu behandeln, z.B. die Häufigkeit des Auftretens von Merkmalsausprägungen auszuweisen, Mittelwerte und Streuungen zu berechnen, Beziehungen zwischen zwei oder mehr Merkmalen herzustellen usw. Für die Lösung jeder dieser Aufgaben benötigte man allerdings jeweils ein eigenes Programm. Durch das Zusammenfassen solcher Programme zu Programmsammlungen versuchte man, diese unkoordinierten Einzellösungen effizienter zu gestalten. Diese Programmsammlungen wurden dann anderen Anwendern zur Verfügung gestellt.
- Danach bemühte man sich, Programme zur Verwaltung dieser Programmsammlungen zu entwickeln, insbesondere mit dem Ziel, den Aufruf der statistischen Auswertungsprogramme über eine einfache (Kommando-) Sprache zu ermöglichen. Für diese Programm'Pakete' bestand ein beträchtlicher Bedarf in der Medizin, den Wirtschaftswissenschaften, der Psycho-

logie und den Sozialwissenschaften. In diesen Programm-Paketen waren eine Vielzahl von (Einzweck-) Programmen zusammengeschürt.

- In der dritten Phase kamen Programme zur Verwaltung der auszuwertenden Datenbestände hinzu. Die Integration von Programmsammlungen und Datenverwaltung unter einer einheitlichen Kommando-Sprache war die Geburtsstunde der statistischen Auswertungssysteme (Statistik-Softwarepakete). Die Vorteile von Programmsystemen (einheitliche Kommando-Sprache, einfach zu erlernende Syntax, leichtes Datenmanagement) gegenüber Programmsammlungen **haben** seit etwa 1970 in den Erfahrungswissenschaften zu einem sprunghaften Anstieg der Datenanalyse mit Hilfe der EDV geführt. Hinzu kam ein immer breiter werdendes Modellspektrum im Rahmen der multivariaten Datenanalyse; die Aufnahme neuerer Verfahren in die gängigen Statistik-Softwarepakete schufen die Voraussetzungen für eine schnelle Verbreitung innovativer Methodenentwicklungen.
- In den 80er-Jahren wurden die meisten Systeme auf die neuen, leistungsfähigen Arbeitsplatzrechner portiert. Das führte zu einer Verbesserung der Mensch-Maschine-Schnittstelle mit zunehmenden interaktiven Möglichkeiten, auch bei der graphischen Darstellung der Auswertungsergebnisse. Die Bedienung der Programmsysteme wurde durch die Entwicklung von Benutzeroberflächen weiter vereinfacht. Sogenannte Menüsteuerungen oder graphische Benutzeroberflächen (vor allem unter Windows) haben den Komfort gegenüber der reinen Stapelverarbeitung (als Abarbeitung von Kommando-Sequenzen) erheblich erhöht.

Neben den vom Großrechner (»Mainframe«) auf den PC übertragenen Programmen haben sich die exklusiv für den Arbeitsplatzrechner zugeschnittenen Statistik-Programmpakete auf dem Softwaremarkt etablieren können. Der Vorteil dieser Pakete besteht insbesondere darin, daß die Programmierung in diesem Fall konsequent auf die PC-Technologie zugeschnitten ist; in der Programmierung wird entsprechend intensiv von den inzwischen erreichten technischen Standards Gebrauch gemacht. Diese Entwicklungen (zum einen Portierungen, zum anderen speziell für den PC entwickelte Statistiksoftware-Pakete) haben heute zu einer fast unüberschaubaren Produktpalette geführt

Um statistische Auswertungssysteme beschreiben zu können, ist es günstig, sich zunächst mit Daten- und Programmbanksystemen zu befassen. Programmpakete enthalten eine Datenbank, in der die Rohdaten gespeichert sind, und eine Programmbank, in der die Programme zu den speziellen Analyseverfahren zu Gruppen (d.h. Prozeduren) zusammengefaßt sind. Diese beiden Komponenten werden aufgebaut und verwaltet durch ein Daten- und Programmbankverwaltungssystem, das wiederum als eine Menge von Programmen implementiert ist. Es kann Programme aus der Programmbank aktivieren und auf Datenstrukturen in der Datenbank zugreifen³. Weitere Aufgaben sind etwa die Sicherung

³ Beispiel: Das Verwaltungssystem aktiviert ein Programm 'Mittelwert' und übergibt

(Archivierung) von Datenstrukturen. Das Verwaltungssystem ist die wichtigste Schnittstelle für den Anwender. Es regelt die Ein- und Ausgabe von Daten und führt üblicherweise ein 'Logbuch', das heißt ein Protokoll aller durchgeführten Aktionen. Die Analyseergebnisse selbst werden wie Daten in die Datenbank abgelegt und können mit Hilfe des Verwaltungssystems angesehen und auch verändert werden. Die Abbildung 1 illustriert noch einmal die Grundidee eines integrierten statistischen Programmsystems.

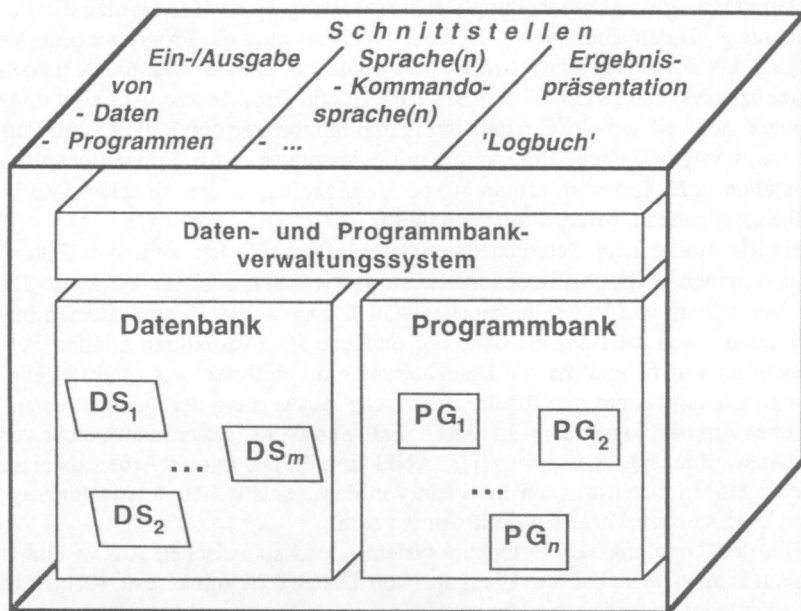


Abb. 1: Architektur eines statistischen Daten- und Programmbanksystems (DS: Datenstruktur; PG: Programmgruppe).

(Quelle: Haag, U./Haux, R./Kieser, M., 1992: Statistische Auswertungssysteme. Stuttgart/Jena/New York: G. Fischer, S. 13).

Statistik-Softwarepakete, die dem Anwender eine inzwischen fast unüberschaubare Zahl statistischer Methoden der Datenanalyse bereitstellen, sind insbesondere in der universitären Forschung zu einem unentbehrlichen Werkzeug der empirischen Erfahrungswissenschaften geworden. Darüberhinaus besitzen zahlreiche Unternehmen und Behörden mindestens eines der großen Statistik-Programmpakete. Diese statistischen Auswertungssysteme verlangen von ihrem Anwender keinerlei Kenntnisse von Programmiersprachen; vielmehr kön-

diesem Programm eine bestimmte Datenstruktur, um den Mittelwert der darin enthaltenen Daten zu einer bestimmten Variablen zu berechnen.

nen auch Anfänger nach Erlernen einer einfachen Kommandosprache Auswertungen vornehmen, die zuvor nur speziell ausgebildetem Personal vorbehalten war.

Wir wollen abschließend noch kurz auf eine spezielle Klasse verwandter Systeme eingehen, den sogenannten Datenbanksystemen, die in den letzten Jahren zunehmend auch in den Erfahrungswissenschaften eingesetzt werden.

Ist die eigentliche Aufgabe des Daten- und Programmbanksystems nur die Datenerfassung und Datenhaltung/Datenverwaltung, so verwendet man die Bezeichnung »Datenbanksystem«. Dementsprechend muß die Programmbank lediglich das Datenbankverwaltungssystem enthalten. Die zur Zeit existierenden Datenbanksysteme (wie z.B. dBASE, Paradox, FoxPro, Access usw.) sind mittlerweile so weit entwickelt, daß ihre Programmbanken zusätzlich zum Verwaltungssystem einfache Programme mit elementaren Statistikoperationen bereitstellen (z.B. sortieren, ermitteln von Häufigkeiten in den einzelnen Merkmalsausprägungen, Mittelwertberechnung).

Häufig findet man Schnittstellen, über die man Daten zwischen Datenbanksystemen und statistischen Auswertungssystemen austauschen kann. Da die Verwaltung von Daten in manchen statistischen Auswertungssystemen immer noch etwas mühselig ist bzw. nur einfache Datenstrukturen erlaubt, verwendet man manchmal für die Datenhaltung ein komfortableres (und flexibleres) Datenbanksystem und für die statistische Auswertung der Daten ein statistisches Auswertungssystem. In diesem Fall werden die Daten unmittelbar vor der Auswertungsphase übertragen (s. Abbildung 2). Die in eine Arbeitsdatei zu übertragenden Daten müssen in einem von dem statistischen Auswertungssystem unterstützten Format erstellt worden sein^{*}.

Bei der Kopplung von Datenbanksystemen und statistischen Auswertungssystemen muß man bei der Übergabe von Daten u.U. mit einem Verlust an Informationen rechnen. Zur Übergabe muß man bei den meisten existierenden Statistik-Programmpaketen jede Datenstruktur in eine sequentielle Datei transformieren (die zudem oft nur Variablen eines reellwertigen Datentyps enthalten darf). Dennoch sei angemerkt, daß die Verwaltung von Daten im Rahmen statistischer Auswertungssysteme nur im begrenzten Ausmaß möglich ist. Flexiblere Lösungen bei komplexen Datenstrukturen garantieren allein die speziell für diese Problemstellungen konzipierten Datenbanksysteme. Sie werden daher zunehmend für die erste Phase der Datenaufbereitung eingesetzt und über Transport-Schnittstellen an statistische Auswertungssysteme zwecks differenzierter statistischer Auswertung übergeben.

Vor jeder Datenerfassung und Datenanalyse muß zunächst aus der Palette der angebotenen statistischen Programmsysteme ein bestimmtes Programm ausgewählt werden. Dies ist insofern eine (Vor-) Entscheidung, weil der gesamte weitere Arbeitsvorgang davon zu einem großen Teil beeinflußt wird. Es ist

*** Der Datentransfer zwischen Datenbanksystemen und Statistik-Programmpaketen ist inzwischen bei nahezu allen Anbietern sehr komfortabel gestaltet.**

jedoch anzumerken, daß zwischen den »großen« Programmpaketen ein direkter Datenaustausch zumindest über sog. »Export«-Dateien möglich ist. Damit lassen sich die mit Hilfe eines speziellen Programmpakets erstellten Arbeitsdateien in weiteren Statistik-Softwaresystemen weiterverarbeiten.

Um dem Anwender die Entscheidung zu erleichtern, werden im folgenden die wichtigsten Statistikprogramme skizziert und vergleichend gegenübergestellt.

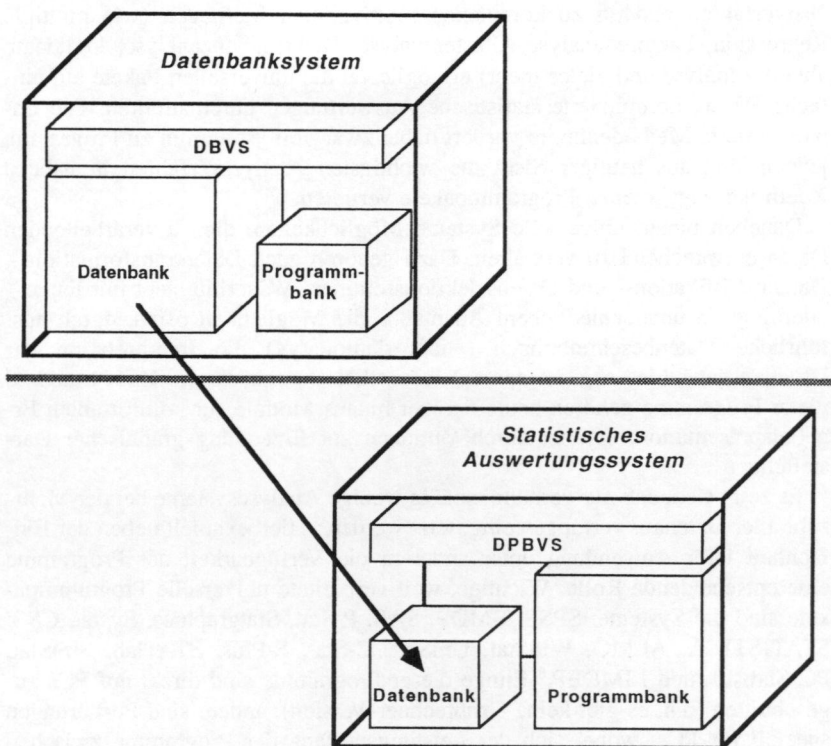


Abb. 2: Kopplung von Datenbanksystem und statistischen Auswertungssystem (DBVS: Datenbankverwaltungssystem; DPBVS: Daten- und Programmbankverwaltungssystem; Pfeil: Datenübergabe vom Datenbanksystem zum statistischen Auswertungssystem).

(Quelle: Haag, U./Haux, R./Kieser, M., 1992: Statistische Auswertungssysteme. Stuttgart/Jena: G. Fischer, S. 32).

1. Eine Klassifikation von Statistikprogrammen

Bei einer Einteilung von Statistikprogrammen nach dem Verwendungszweck wird zwischen *universellen* Programmpaketen und Systemen für *spezielle Anwendungen* unterschieden. Universelle Statistik-Programmpakete beinhalten eine Vielzahl von statistischen Methoden, die für die unterschiedlichsten Problemstellungen eingesetzt werden können. Angefangen bei einfachen Häufigkeitsverteilungen und deskriptiven Statistiken, unterschiedlichen Schätz- und Testverfahren bis hin zu komplexen multivariaten Methoden (wie multiple Regression, Faktorenanalyse, Clusteranalyse, Diskriminanzanalyse, loglineare Tabellenanalyse und vieles mehr) ermöglichen die universellen Pakete auf einfache Weise, komplizierte statistische Auswertungen durchzuführen. Die implementierte Methodenbreite variiert dabei zwar von Programm zu Programm, jedoch sind aus heutiger Sicht die wichtigsten Analyseverfahren in nahezu jedem der angebotenen Programmpakete vertreten.

Daneben bieten universelle Systeme Möglichkeiten, die zu verarbeitenden Daten entsprechend zu verwalten. Dazu gehören auch Datentransformations-, Datenmodifikations- und Datenselektionsroutinen. Weiterhin steht mit ihnen - allerdings in unterschiedlichem Ausmaß - die Möglichkeit offen, durch ausführliche Datenbeschreibungen (»data dictionary«) die Interpretation der Druckausgabe (des sog. »Outputs«) wesentlich zu erleichtern. Zum Standard vieler Programme gehören heute darüber hinaus Module zur komfortablen Ergebnispräsentation, ergänzt durch Optionen zur Erzeugung graphischer Darstellungen.

Es zeigt sich, daß nur verhältnismäßig wenige Analysesysteme bei der Mehrzahl aller Datenauswertungen eingesetzt werden.⁵ Hierbei spielt neben der Einfachheit ihrer Anwendung unter anderem die Verfügbarkeit der Programme eine entscheidende Rolle. Wichtige, weit verbreitete universelle Programmpakete sind die Systeme: SPSS, BMDP, SAS, P-Stat, Statgraphics, Systat, CSS: STATISTICA, ALMO, WinStat, UniStat, C-Stat, S-Plus, STATlab, SimStat, PC-Statistik und LIMDEP.⁶ Einige dieser Programme sind direkt auf PCs zugeschnitten (d.h. es gibt keine Großrechner-Version), andere sind Portierungen speziell für PCs, wobei sich der Leistungsumfang der Programme zwischen Großrechner und PC inzwischen kaum mehr unterscheidet.

Neben den universellen Statistik-Programmpaketen, die Methoden für eine Vielzahl von Anwendungsbereichen enthalten, gibt es Programme, die für spezielle methodische Fragestellungen konzipiert sind und deshalb ganz bestimmte statistische Verfahren umfassen. Als »Spezialisten« unter den Statistik-Softwareprogrammen schließen sie die Lücken im Leistungsumfang der univer-

⁵ Vgl. die Untersuchung von Bausch, T./Bankhofer, U.: Statistical Software packages for PCs - a market survey, in: Statistical Papers 33 (1992), S. 283 - 306

⁶ In der Auswahlbibliography sind die Einführungen in die Programmsysteme jeweils gesondert aufgenommen.

seilen Programmpakete. In der folgenden Skizzierung dieser Programme wurde besonders auf die Verfügbarkeit der Programme für PCs unter dem Betriebssystem DOS geachtet.

Für *Zeitreihen* wurde z.B. eine wichtige Gruppe von Spezialprogrammen konzipiert. Diese Programme schließen häufig Kurvenanpassungen, Glättungsverfahren, Spektralanalyse, Box-Jenkins-Analyse, Methoden zur Schätzung von Parametern in ökonometrischen Ein- und Mehrgleichungsmodellen sowie die Prognose von Zeitreihenwerten ein. Als weit verbreitete Vertreter dieser Gattung von Statistik-Spezialprogrammen seien *MicroTSP* (»Time Series Processor«), *RATS* (»Regression Analysis of Time Series«), *AUTOBOX*, das unter SYSTAT (ausführlich s.u. im Rahmen der Besprechung der universellen Programmpakete) vertriebene Zeitreihenanalysemodul *MESOSAUR* und das SPSS-Modul *TRENDS* genannt, die auch für PCs verfügbar sind. Für multiple Zeitreihenanalysen eignet sich das in jüngster Zeit entwickelte Programm *MULTI*, das an der Universität Kiel am Institut für Statistik und Ökonometrie entwickelt wurde. Weitere Spezialprogramme sind *STAMP* (»Structural Time Series Analyser, Modeller and Predictor«) und *SCA* (»Scientific Computing Associates«)⁷.

Bezugsquellen:

- RATS: VAR Econometrics, 134 Prospect Ave. South, Minneapolis MN 55419, USA;
- MicroTSP 7.0: GEWUS, Gardeschützenweg 90, 12203 Berlin oder: Quantitative Micro Software, 4521 Campus Drive, Suite 336, Irvine, California 92715, U.S.A.;
- AUTOBOX: Automatic Forecasting Systems (AFS) Inc., P.O. Box 563/Hatboro, Pennsylvania 19040, U.S.A.;
- SCA Statistical System (Forecasting and Time Series Analysis): Statistical Computing Associates Corp., P.O. Box 4692, Oak Brook, Illinois 60522, U.S.A.
- MULTI: Institut für Statistik und Ökonometrie, Universität Kiel, Olshausenstr. 40, 24118 Kiel, Germany.
- MESOSAUR 1.0: STATCON B. Schäfer, Mündener Str. 1, 37213 Witzenhäusen, Germany;
- SCA: Scientific Computing Associates Corp., 913 W. Van Buren Street, Suite 3H, Chicago, Illinois 60607-3528, U.S.A.;
- STAMP: Prof. A.C. Harvey, Department of Statistical and Mathematical Sciences, London School of Economics and Political Sciences, Houghton Street, London WC2AE, U.K., Tel.: 01-955-7633.

⁷ Eine ausführliche Besprechung dieser Spezialprogramme findet der interessierte Leser in: Rahlf, Th., 1994: PC-Programme zur Zeitreihenanalyse: Datenmanagement, Graphik und univariate Analyseverfahren. SPSS, Systat, STATISTICA, MicroTSP, MESOSAUR, in: *Historical Social Research/Historische Sozialforschung*, 19/3, S. 78-123. Ein weiterer Beitrag zu dem Programm SCA von Th. Rahlf erscheint demnächst in der HSR.

Programme zur Schätzung von Parametern in linearen Strukturgleichungsmodellen (Pfadanalyse) gewinnen in den Wirtschafts- und Sozialwissenschaften immer mehr an Bedeutung. Neben der Schätzung von Mehrgleichungsmodellen mit beobachteten Variablen bieten sie die Möglichkeit, beobachtete Indikatoren zu latenten Variablen zu verknüpfen, deren hypothetische Kausalstruktur wiederum simultan geschätzt wird. Dem Ansatz der Faktorenanalyse folgend, werden die latenten Variablen zunächst durch direkt gemessene Indikatoren spezifiziert. Die Beziehungen zwischen den latenten Variablen werden, analog zum pfadanalytischen Kausal-Ansatz, in ein Strukturmodell übersetzt, wobei hypothetische Null-Restriktionen in den Regressionsgleichungen der latenten Variablen angenommen werden können. Das bekannteste System dieser Programmattung ist *LISREL* (»Linear Structural Relation System«, zur Zeit in der Version 8), das auch in Anbindung an SPSS-X betrieben werden kann. Auch dieses Programm gibt es als PC-Version unter DOS oder wahlweise unter Windows (Europäischer Vertrieb: ProGAMMA, P.O. Box 841, 9700 AV Groningen, The Netherlands).

Verfahren der numerischen Klassifikation teilen Objekte, anhand der für ein Untersuchungsziel wichtig erscheinenden Merkmale, in möglichst homogene Klassen ein. Ein Spezialprogramm nur für Clusteranalysen ist z.B. das System *CLUSTAN* von D. Wishart. Das System bietet sowohl für metrische als auch für binäre Daten eine große Auswahl von Ähnlichkeitsmaßen an (z.B. allein 15 unterschiedliche Hauptverfahren zur Datenklassifikation und unterschiedliche Verfahren zur Überprüfung empirisch gewonnener Klassifikationen). Das Programm enthält ferner ein Modul zur Dateneingabe und Datenaufbereitung. Die Behandlung von fehlenden Werten und Labels ist möglich. Die Ergebnisse der einzelnen Klassifikationen können sowohl tabellarisch als auch graphisch (auf Bildschirm oder Postscript-Druckern) dargestellt werden. Neu ist vor allem die verbesserte Bildschirmdarstellung (Quelle: D. Wishart, General Sales Manager, London Univ. Computing Services Ltd., 39 Gordon Square, London W.C. 1; Preis für die Einzelkopie: US \$ 1000,- ; Anfragen können auch gerichtet werden an Cornelia Züll, ZUMA, Zentrum für Umfragen, Methoden und Analysen, B2, 1, 68072 Mannheim, Tel.: 0621/18004-0).

Ein Programm zur mehrdimensionalen Kontingenztabellenanalyse nach dem sog. »Weighted Least Squares« - Ansatz (das von Grizzle, Starmer und Koch vorgeschlagene GSK-Modell) zur multivariaten Analyse diskreter Daten ist *NONMET II* von H. M. Kritzer. Das Modell läßt sich kurz als Anwendung der gewichteten multiplen Regression mit Dummy-Variablen auf mehrdimensionale Kontingenztabellen charakterisieren (Quelle: Department of Political Science, Univ. of Wisconsin, Madison, Wisconsin 53706; Vertrieb in Deutschland: ZUMA, Mannheim, Anschrift s.o.).

Bei der Analyse nicht-metrischer Daten findet in der jüngeren Literatur zunehmend ein alternativer Ansatz Verwendung: die Korrespondenzanalyse. Sie dient ebenfalls der multivariaten Analyse von Kontingenztabellen. Dabei han-

delt es sich aber um ein symmetrisches Verfahren (ohne expliziter abhängiger Variable), das große Ähnlichkeit mit der Hauptkomponentenanalyse besitzt. Für praktische Anwendungen wurde für den PC das Programm *SIMCA* von M. Greenacre entwickelt (Quelle: M.J. Greenacre, Department of Statistics, University of South Africa, P.O. Box 392, Pretoria 0001, South Africa; Auskunft erteilt Jörg Blasius, Zentralarchiv für empirische Sozialforschung an der Univ. zu Köln, Postfach 410960, 50869 Köln, Tel.: 0221-47694-0).

Letztlich bleibt anzumerken, daß zunehmend Programmentwicklungen für die unterschiedlichsten Spezialgebiete der Statistik angeboten werden. Um einige Beispiele zu nennen: *LCA* zur diskreten Latent-Class Analyse, *UTA* zur Latent-Trait Analyse nach dem RASCH-Modell (vertrieben von R. Langeheine, Univ. Kiel, Institut für die Pädagogik der Naturwissenschaften) und das inzwischen weit verbreitete Programm *TDA* (»Time Dependent Analysis«), das speziell auf die Probleme der Ereignisdatenanalyse zugeschnitten ist. Dieses Programm wird von Götz Rohwer mit einer ausgezeichneten Programmbeschreibung gegen eine geringe Schutzgebühr vertrieben (Universität Bremen, FB 8, Universitätsallee, 28359 Bremen, Postfach 330440).

Häufig besteht die Schwierigkeit allerdings darin, die Existenz solcher Programme überhaupt ausfindig zu machen. Einen sehr guten Überblick zur Statistiksoftware-Entwicklung geben die regelmäßig veröffentlichten Sammelbände *SoftStat*, die von F. Faulbaum herausgegeben werden (s. den Abschnitt 1 in der Auswahlbibliographie). Sie dokumentieren die im Turnus von 2 Jahren stattfindenden Statistiksoftware-Konferenzen der ZUMA.

Um Mißverständnisse zu vermeiden, sei noch angemerkt, daß die universellen Programmpakete auch Implementationen spezieller Methoden enthalten. Viele Verfahren der SpezialSoftware sind daher auch in den o.g. Programmsystemen verfügbar. Aufgrund ihres Zuschnitts auf sehr spezifische Problemstellungen ist die SpezialSoftware allerdings meist den Implementationen in den universellen Programmen überlegen. Hinzu kommt, daß die »Spezialisten« meist auch die jüngeren Weiterentwicklungen auf dem Gebiet spezieller Datenanalyseverfahren berücksichtigen. Die Verwendung von (Einzweck-)Programmen bleibt somit dort notwendig und sinnvoll, wo neue Verfahren der Datenanalyse entwickelt bzw. weiterentwickelt werden und wo universelle Statistik-Programmpakete Lücken aufweisen oder in ihren Datenanalyse-Unterprogrammen für bestimmte Fragestellungen nicht differenziert genug sind.

2. Vorteile und Gefahren der Anwendung von Statistik-Programmpaketen

Die praktische Anwendung von statistischen Verfahren wird in starkem Maße von der EDV beeinflusst. Dieses Hilfsmittel hat der Statistik einerseits außerordentliche Möglichkeiten eröffnet, droht aber andererseits, die Anwendung

von statistischen Verfahren in Mißkredit zu bringen*. Diese beiden Aspekte sollen im folgenden kurz erläutert werden.

Die elektronische Datenverarbeitung erleichtert die schnelle und zuverlässige Aufbereitung großer Datenmengen. Darüberhinaus hat das Aufkommen von Computern die Auswertung von Daten mit Hilfe von statistischen Methoden in einer früher kaum vorstellbaren Weise vereinfacht. Die weltweite Verfügbarkeit umfangreicher und nutzerfreundlicher Statistikpakete für alle Leistungsklassen der Rechentechnik hat wesentlich dazu beigetragen, die Anwendung auch komplizierter Methoden der mathematischen Statistik in den empirischen Wissenschaften zu fördern. Besonders die universellen Programmpakete sind ohne Informatik- und Programmierkenntnisse anwendbar. Meist genügt ein kurzer Einführungskurs, um Auswertungen durchführen zu können. Dem Anwender steht dann eine große Palette statistischer Methoden zur Verfügung, so daß ohne die Hilfe eines Spezialisten selbst die anspruchvollsten Analyseverfahren möglich werden. Die Entwicklung der PC-Statistiksoftware in den letzten Jahren zeigt darüber hinaus einen deutlichen Trend zu komplett menügesteuerten Benutzeroberflächen, die dem Anwender selbst die Schreibarbeit für die programmspezifischen Systemkommandos abnehmen. Diese erkennbaren Vorteile von Programmsystemen haben seit den 70er Jahren in den Erfahrungswissenschaften zu einem enormen Anstieg der Datenanalyse mit Hilfe der EDV geführt. Diese Entwicklung ist noch längst nicht abgeschlossen und erfährt mit der Herausgabe überarbeiteter oder gänzlich neuer Versionen sowie ihren Implementationen auf Mikrocomputern ständig neue Impulse. Die Verwendung von Einzelprogrammen ist dagegen im Vergleich stark zurückgedrängt worden.

Das Bild wäre unvollständig, wenn hier nicht auch die Nachteile erwähnt würden, die aus dem Mißbrauch der EDV entstehen. Schwerwiegend sind die Gefahren, die der unkritische Einsatz des Hilfsmittels EDV (bzw. eines Statistik-Programmpakets) für den statistisch unzureichend ausgebildeten Benutzer mit sich bringt, der Computerprogramme als Surrogat für Kenntnisse der statistischen Methodenlehre ansieht. Das »Verführungspotential« von Programmpaketen läßt vielfach Benutzer ohne hinreichende Statistikenkenntnisse einfach »losrechnen«. Die Sorglosigkeit im Umgang mit Rechenprogrammen für statistische Methoden zeigt sich oftmals auch darin, die erhobenen Daten nach möglichst vielen Verfahren auswerten zu lassen. Bezüglich der dabei erzielten Resultate bestehen aber erhebliche Zweifel, ob sie statistisch sinnvoll sind, d.h. die untersuchte Fragestellung adäquat beantworten. Die Vielfalt statistischer Prozeduren lädt zu einer leichtfertigen Anwendung ein: oft wird nicht bedacht, welche Voraussetzungen die erhobenen Daten aufweisen müssen, um bestimmte statistische Verfahren anwenden zu können. Die Interpretation der Ergebnisse erfolgt dann rezeptartig anhand einschlägiger Statistik-Handbücher. Eine

* S. insbesondere Neider, J.A., 1977: Intelligent programs, the next stage in statistical computing, in: Barra et al. (eds.), 1977: Recent developments in statistics. Amsterdam: North-Holland.

solche Vorgehensweise führt zwangsläufig zur Produktion von Forschungsartefakten. Zum anderen erleichtert die zunehmend vereinfachte Programmsteuerung ein unreflektiertes Durchspielen aller möglichen Verfahren, was zwangsläufig zu stereotypen Auswertungsmustern führen kann.

Versucht man, die Fehlerursachen bei der Anwendung statistischer Verfahren zu ordnen, so lassen sich die häufigsten Fehlerursachen in folgende Kategorien einteilen:

- Fehler durch die Auswahl falscher Methoden bzw. Nichtanwendung der richtigen Methoden für eine gegebene Problemstellung (Beschränkung des Methodenspektrums),
- Fehler in der Anwendung von Methoden, d.h. Berechnung falscher Statistiken (Parameter) oder Wahl von unzulässigen Programmoptionen,
- Anwendung von Methoden, deren Voraussetzungen (z.B. Verteilungsannahmen, Stichprobenumfang) nicht gegeben sind oder die wegen Anomalien im Datenmaterial (z.B. Autokorrelation bei Zeitreihen, Multikollinearität, Nichtlinearität der Variablenbeziehungen etc.) nicht angewendet dürfen,
- Fehler 3. Art, d.h. eine exakte Lösung des falschen Problems,
- fehlerhafte Interpretation der Ergebnisprotokolle von Auswertungen, z.B. wegen einer unklaren Darstellung der Ergebnisse oder durch eine unzulässige Beschränkung der Ergebnisdarstellung.

Wie steht es mit der Befürchtung des renommierten englischen Statistikers John A. Neiders, die Statistik gerate aufgrund der kritiklosen Anwendung statistischer Programme in Mißkredit?

Sicherlich sind die meisten der verbreiteten statistischen Auswertungssysteme nicht in der Lage, den Anwender konstruktiv bei der statistischen Datenanalyse zu kontrollieren. Die Wege zu einer Verbesserung dieses Zustandes sind jedoch bekannt: sorgfältige und methodische Entwicklung der Programme und Dokumentation der Lösungsalgorithmen, Verbesserung der Datenbankverwaltung, Integration von Wissen über Daten und Programme. Den ersten dieser drei Wege hat man bereits beschritten, den zweiten hat man weitgehend erforscht, auf dem dritten sind die ersten Schritte in Richtung sogenannter »Wissensbasierter Systeme (oder Expertensysteme)« getan. Wir gehen im letzten Abschnitt gesondert auf diese Entwicklung ein. In diesem Zusammenhang sei hier bereits angemerkt, daß statistische Expertensysteme zur Zeit noch in der praktischen Erprobung sind.

In den folgenden Abschnitten wollen wir die wichtigsten Statistik-Programmpakete kurz vorstellen und anhand ausgewählter Leistungsmerkmale vergleichen. Aus der wachsenden Zahl der Softwareprodukte wurden diejenigen ausgewählt, die als Paket mit allgemeinem Methodenspektrum zu bezeichnen sind. Außerdem sollten die Pakete auf IBM-kompatiblen PCs lauffähig und in Deutschland verfügbar sein. Besonderes Gewicht wurde bei der Gegenüber-

Stellung auf das vorhandene Methodenspektrum, die Bedienerführung und die Möglichkeiten zur graphischen Ergebnispräsentation gerichtet.

3. Übersicht zu den Statistik-Programmpaketen

Die Suche nach der angemessenen computerunterstützten Datenanalyse für die jeweilige Fragestellung erweist sich für viele Wissenschaftler noch immer als mit Hürden und Hindernissen gespickt. Es fehlt oft der Überblick, welche oder welches der statistischen Auswertungssysteme den eigenen Daten und konkreten Fragestellungen oder Anforderungen (z.B. im Hinblick auf die graphische Gestaltung von Ergebnissen) am ehesten entsprechen. So stehen für die statistische Auswertung von Daten für alle Computertypen, seien sie nun groß (»mainframe«), mittel (»Workstation«) oder klein (»personal Computer«), eine Vielzahl mehr oder minder mächtiger Analysesysteme zur Verfügung. Die Unübersichtlichkeit des Marktes für statistische Programmpakete mit unterschiedlichen Versionen und Ausgaben der einzelnen Programmsysteme für verschiedene Computertypen und den dazugehörigen Betriebssystemen sollte aber niemanden von computerunterstützter Datenanalyse zurückschrecken lassen.

Die in den vergangenen Jahren enorm angestiegene Zahl von Statistikanwendungen auf dem PC weisen auch in eine deutliche Richtung. Diese Entwicklung wurde vornehmlich durch die sinkenden Hard- und Softwarekosten und die gewaltig zugenommene Leistungsfähigkeit der PCs verursacht. Die Verarbeitung großer Datensätze mit akzeptablen Laufzeiten war lange Zeit eine Domäne der Großrechner. In den vergangenen Jahren sind aber Software-Programme für den PC auf den Markt gekommen, die sich in ihren Leistungsmerkmalen von der etablierten Großrechner-Software kaum noch unterscheiden. Die PC-Software hat in Einzelfällen sogar fortgeschrittene Entwicklungen implementiert, die auf den Großrechner-Versionen fehlen.

In diesem Abschnitt soll ein allgemeiner Überblick über die bekanntesten und am weitesten verbreiteten statistischen Software-Pakete für den PC gegeben und Kriterien für die Auswahl und Beurteilung von statistischen Datenanalysesystemen aufgezeigt werden. Informationen über diese Programmpakete veralten sehr schnell, da diese Produkte einer ständigen Weiterentwicklung unterliegen. So sind z.B. viele Angaben in dem lesenswerten Aufsatz von T. Bausch/U. Bankhofer, 1992: Statistical software package for PCs - a market survey (in: Statistical Papers, 33, S. 283-306) über statistische Software-Programme bereits überholt, weil fast alle Hersteller heute gegenüber den dort getesteten Produkten weiterentwickelte Versionen anbieten. Insofern kann dieser Beitrag auch nur den Stand der derzeitigen Entwicklung reflektieren und soll einen ersten Anstoß für die intensivere Auseinandersetzung mit einzelnen Produkten geben.

Je nach Datenmaterial, Analyseziel und Vorkenntnissen des Nutzers variieren die Anforderungen an Statistikpakete sehr stark, so daß sich jeder Anwender sein eigenes Anforderungsprofil entwickeln muß. Die Wahl eines Datenanalysesystems hängt - neben seiner schlichten Verfügbarkeit - selbstverständlich auch vom individuellen Arbeitsstil und den damit verbundenen Vorlieben ab. Auch in dieser Hinsicht erscheint es wichtig, auf Alternativen hinzuweisen. In der nachfolgenden Beschreibung der Einzelprogramme kann zunächst nur sehr allgemein auf die folgenden Kriterien eingegangen werden:

Hardwarevoraussetzungen: Da die Systeme erheblichen Festplattenspeicher zur Implementation benötigen, ist in jedem Fall eine Festplatte von mindestens 20 MB erforderlich. Einige Statistik-Programme sind modular aufgebaut, d.h. der Benutzer kann, um Speicherplatz (und Kosten) zu sparen, auf Teile des Gesamtsystems und damit auf ständig verfügbare Leistungsfähigkeit verzichten. Als weitere Hardwarevoraussetzung ist ein Hauptspeicher von 640 KB für DOS-Versionen empfehlenswert, für Statistikprogramme, die unter der graphischen Benutzeroberfläche Windows implementiert sind, mindestens 2 MB. Ferner ist ein mathematischer Coprozessor für 386er Rechner erforderlich, da sich sonst das Laufzeitverhalten drastisch verschlechtert.

Benutzerführung: Viele Programmpakete sind interaktiv nutzbar und verfügen über eine Menüsteuerung. Ein Menü ist nichts anderes als ein hierarchisch geordneter Katalog möglicher Befehle, die dem Benutzer zur Auswahl stehen. Die Stärke der PC-Anwendung liegt in der Dialogverarbeitung, d.h. jeder Befehl wird direkt ausgeführt. Diese Möglichkeit wird von den Datenanalysesystemen in mehr oder weniger großem Umfang unterstützt. Neben der menügesteuerten Dialogverarbeitung ermöglichen einige Datenanalysesysteme auch auf dem PC den Batch- oder Stapelverarbeitungsbetrieb. Darunter ist zu verstehen, daß zunächst ein Programm geschrieben wird, welches einen Stapel von Befehlen enthält, der an das jeweilige Datenanalysesystem übergeben wird. Der Befehlsstapel wird dann abgearbeitet, als ob die Befehle über Tastatur oder Menüabruf nacheinander direkt eingegeben würden. Die Befehle sind in der jeweiligen Kommandosyntax zu schreiben, deren Schwierigkeitsgrad je nach Programmpaket sehr unterschiedlich ausfällt. Einige Programme verfügen zusätzlich noch über eine eigene Programmiersprache. Mit Hilfe dieser Sprache lassen sich, da z.B. Matrizenoperationen und Rückgriffe auf statistische und mathematische Routinen zur Verfügung stehen, eigene (auf individuelle Datenanalyseprobleme zugeschnittene) Auswertungsroutinen realisieren und als sogenannte Makros in das Gesamtsystem einbinden.

Hilfe-Menü: Zur Benutzerfreundlichkeit der Programme trägt die Möglichkeit bei, während einer Dialogsitzung direkt Hilfestellungen anzufordern. Die abrufbaren Informationen sind teilweise recht ausführlich. Bei manchen Programmsystemen kann allerdings auf den Gebrauch des Handbuchs nicht verzichtet werden, will man deren Möglichkeiten ausschöpfen. Einige Programmsysteme sind darüberhinaus vollständig über deutschsprachige Menüs steuerbar.

Output-Gestaltung: Für den Anwender ist von großer Bedeutung, daß der Ergebnis-Output eines Programms übersichtlich und leicht lesbar ist. Die Datenanalyseprogramme erfüllen weitestgehend diese Anforderungen. Eine Vielzahl von Programmen erlauben eine spezielle Gestaltung des Outputs, wodurch sich die Möglichkeit ergibt, diesen ohne aufwendige Bearbeitung in einen Berichtstext zu kopieren. Ferner ist es von Bedeutung, inwieweit die Systeme graphische Darstellungsmöglichkeiten erlauben. Einige Programmsysteme verfügen über ein integriertes Graphikmodul, das die Ausgabe sogenannter Präsentations- oder Businessgraphiken in hoher Qualität erlaubt (sog. hochauflösende Graphiken). Einige Programme stellen den Benutzern lediglich eine Schnittstelle zur Verfügung, mittel derer die Daten- bzw. Ergebnisübergabe zu einem eigenständigen Graphikprogramm (z.B. Harvard Graphics) möglich ist.

Daten- und Informationsaustausch: Bei der Auswahl der Statistik-Software sollte nicht ausschließlich ihr Leistungsspektrum betrachtet werden, sondern ein besonderes Augenmerk sollte auch auf die Möglichkeit der Kommunikation mit anderen Programmen gelegt werden. Dies betrifft besonders die Möglichkeit sogenannte Systemfiles auszutauschen, in denen neben den Rohdaten auch Daten- und Dateibeschreibungen in binärer, vom jeweiligen Programmpaket direkt lesbarer Form gespeichert sind. Alle Datenanalyzesysteme unterstützen selbstverständlich die Möglichkeit, Dateien im ASCII-Format auszutauschen. Dabei gehen aber möglicherweise eine Vielzahl von Informationen, wie z.B. Variablen- und Ausprägungsetiketten oder die Definition fehlender Werte, verloren. Einige Programmpakete stellen zum Datentransfer eine DIF-Schnittstelle («Data Interchange Format») zur Verfügung. Dieses Format ist zu einer Quasi-Norm geworden, Rohdaten und Datenbeschreibungen auszutauschen, und wird über Hardware- und Betriebssystemgrenzen hinweg von vielen Programmen unterstützt. Auf eine mögliche Ergänzung zu den Datenanalysesystemen sei in diesem Zusammenhang hingewiesen. Das Spezialprogramm »DBMS/COPY« ermöglicht den direkten Filetransfer zwischen einer Vielzahl von Datenanalysesystemen und Datenbank-Programmen auf DOS-Rechnern⁹. Das Programm erlaubt, Files der Datenanalyzesysteme in ein anderes Format zu übersetzen.

Kosten-INutzen-Gesichtspunkt: Das Preis-Leistungsverhältnis spielt sicherlich bei dem universitären Einsatz von Statistik-Programmsystemen eine untergeordnete Rolle (sieht man einmal von der Frage der Verfügbarkeit ab). Hinzu kommt, daß einige Anbieter von Analysesystemen Campus-Lizenzen vergeben, die deutlich unter dem Normalpreis liegen (in der Regel unter 100,— DM). Falls kein Zugang zu einem Datenanalysesystem besteht, soll im Hin-

⁹ Siehe auch den Forschungsbeitrag: Ritter, H./Züll, C, 1992: Erfahrungen und Problemlösungen beim Datenaustausch zwischen Statistikprogrammsystemen, in: ZUMA Nachrichten, 30, S. 109-122 (Zentrum für Umfragen, Methoden und Analysen, Mannheim). Die Autoren berichten ausführlich über ihre Erfahrungen mit dem Programm DBMS/COPY beim Datenaustausch zwischen Statistikprogrammen. Bezugsquelle: DBMS/COPY: Conceptual Software Inc., P.O. Box 56627 Houston, TX 77256, USA.

blick auf finanzielle Restriktionen im Rahmen unserer Besprechung auf preisgünstige Alternativen zu den 'mächtigen' Programmsystemen hingewiesen werden. Sollen nur einmalig relativ geringe Datenmengen für ein kleines Forschungsvorhaben analysiert werden, so ist möglicherweise ein Programm zu einem Preis von unter 500,- DM völlig ausreichend, in Einzelfällen kann selbst ein Public-Domain oder Shareware-Programm für eine eingeschränkte Fragestellung ausreichend sein (vorausgesetzt, das Programm ist gut dokumentiert). Ein leistungsfähiges und bekanntes Shareware-Programm ist STATMA-TE/PLUS. Ausgefeiltere Lösungen bietet das kostengünstige Programmpaket NCSS oder das von ZUMA vertriebene Programm NSDstat. Ein sehr gutes und preisgünstiges Datenanalysesystem ist das menügesteuerte Programm SIM-STAT, das auch über eine gute integrierte Graphik verfügt¹⁰ Trotz der schnellen Entwicklung neuer Produkte für die PC-Welt, ist die 'professionelle' Datenauswertung bislang auf die klassischen Datenanalysesysteme, wie die oben bereits aufgelisteten, angewiesen.

Im Mittelpunkt stehen im folgenden zunächst die allgemeinen Charakteristika einer Auswahl von universellen Programmpaketen. Angesichts der großen Vielfalt von Systemen für die rechnergestützte Datenanalyse ist es eine nahezu unlösbare Aufgabe, umfassende und gleichzeitig detaillierte Beschreibungen der Systemkomponenten zu geben. Wir beschränken uns im folgenden auf die Gesichtspunkte

- Leistungs Spektrum,
- Erlernbarkeit und Anwendbarkeit,
- graphische Darstellungs- und Analysemöglichkeiten und
- Preis-/Leistungsverhältnis.

Der Funktionsumfang wird im nachfolgenden Abschnitt ausführlich tabellarisch dargestellt

SPSS-X, SPSS/PC + und SPSS für Windows

Das wohl am meisten verbreitete und am häufigsten verwandte statistische Programmpaket dürfte das »Statistical Product and Service Solutions«, kurz: SPSS sein (früher »Statistical Package for the Social Sciences«). Das Programmsystem zur Datenverwaltung und statistischen Datenanalyse hat sich seit seinem Erscheinen 1970 in aller Welt bewährt. Es ist mittlerweile an nahezu allen Hochschulrechenzentren sowie in zahlreichen Behörden und Wirtschafts-

¹⁰ Das Programm kostet 65 US-\$. Ferner können Zusatzmodule erworben werden: Erstellung von Präsentationstabellen (\$ 15), Spreadsheet Editor (SIMEDIT, \$ 8 US); Campus Lizenzen können ebenfalls erworben werden. Quelle: Normand Peladeau, 5000 Adam Street, Montreal, QC Canada, H1V 1W5. Alternativ von CompuServe über Internet: 71355.47@compuserve.com; oder: Public Software Library, P.O. 35705, Houston, TX 77235-5705.

unternehmen installiert, für viele verschiedene Rechner und Betriebssysteme verfügbar, sowohl für Großrechner als auch für Personal Computer (incl. Macintosh).

Das Programmsystem ist 1983 als völlig überarbeitete Version SPSS-X (oder SPSS-Extended) vorgestellt worden. Es besteht aus einem Basissystem und mehreren Erweiterungsmodulen, die je nach Bedarf gekauft bzw. für Großrechner gemietet werden können:

MODUL:	LEISTUNGSUMFANG:
Base	Basissystem
Statistics	Statistikverfahren
Advanced Statistics	Zusätzliche Statistikmethoden
Table	Erstellung präsentationsfähiger Tabellen
Trends	Zeitreihenanalysen und Prognosen
Categories	Conjoint- und Korrespondenzanalyse
Graphics	Graphik-Schnittstelle zu MS-CHART
Mapping	Thematische Karten (nur für SPSS/PC+)
Data Entry II	Datenerfassung

Das System wird mit Hilfe einer prozeduralen Kommandosprache gesteuert, die leicht zu erlernen ist. Die PC-Version SPSS/PC+ unterscheidet sich dabei jedoch hinsichtlich ihrer Benutzeroberfläche von ihrem Großrechnerpendant. Eine Menüoberfläche unterstützt den Anwender bei der Zusammenstellung seiner Kommando-Befehlssequenzen. Bei dem Start des Programms erscheinen auf dem Bildschirm drei, sich teilweise überdeckende Fenster. In einem Auswahlfenster kann eine Option gewählt werden, die in einem weiteren Fenster immer kurz erläutert wird. In dem dritten Fenster, das zugleich als Kommandoeditor dient, wird aus den gewählten Optionen die SPSS-Kommando-sequenz generiert. Der so erstellte Batch-Job kann anschließend über eine Funktionstaste gestartet werden. In einem speziellen Menü werden die Befehle ausführlich erläutert. Diese Art der Benutzerführung, die in anderen Programmpaketen in der jeweiligen PC-Version ebenfalls in ähnlicher Form existiert, ist bei diesem Programmpaket am besten verwirklicht. Ersatzweise kann entweder ein sog. Review-Editor benutzt werden, von dem aus die Befehle direkt ausgeführt werden. Schließlich besteht noch die Möglichkeit die Batch-Datei mit einem externen Editor zu erzeugen.

Sämtliche Mitteilungen des Programms und die Analyseergebnisse werden sowohl auf dem Bildschirm angezeigt als auch in einer ASCII-Datei protokolliert. Der Benutzer kann die Ergebnis-Datei entweder mit Hilfe des SPSS-Editors bearbeiten oder in ein Textverarbeitungsprogramm laden.

Das Datenmanagement von SPSS/PC+ ist im allgemeinen sehr gut. Neben einem internen Format können problemlos ASCII-, dBASE- oder Lotus-Dateien eingelesen werden. Mehrere Dateien lassen sich problemlos auf drei Arten verknüpfen: Fallweise, variablenweise und variablenweise nach Übereinstimmung in einem Schlüsselfeld.

Sehr nachteilig ist, daß SPSS/PC+ noch nicht über eine Möglichkeit verfügt, Daten interaktiv, etwa in Form eines Fall- bzw. Variablen-orientierten Spreadsheets zu bearbeiten. Hier besteht als Alternative innerhalb des Programmsystems nur der (zusätzliche) Erwerb des (altertümlichen) interaktiven Datenbearbeitungsmoduls »Data Entry II«. Eine abgespeckte Version ist als »Quick Editor« seit der Version 5 im Basismodul enthalten.

SPSS bzw. SPSS/PC+ ist ein sehr mächtiges und damit auch sehr umfangreiches Programmsystem. Die Beschreibung aller Möglichkeiten, die SPSS bietet, würde den Rahmen dieses Überblicks sprengen. Wir stellen die Analyse-schwerpunkte der einzelnen Statistik-Module daher lediglich in Form einer Übersicht vor.

SPSS Base System: Das Base System bietet eine komfortable Datenverwaltung, einen Editor, grundlegende Auswertungsverfahren der univariaten Datenanalyse, Mittelwertvergleiche, Korrelation, ANOVA-Modelle, Regression, Kreuztabellen, nichtparametrische Tests und Mehrfachantworten-Analysen.

SPSS Professional Statistics: Dieses Modul umfaßt erweiterte Methoden zur Analyse von Beziehungen innerhalb der Daten, wie z.B. Faktorenanalyse, Clusteranalyse, Diskriminanzanalyse und multidimensionale Skalierung.

SPSS Advanced Statistics: Das Modul bietet anspruchsvolle statistische Modelle zur Analyse von Zusammenhängen zwischen Variablen. Die Verfahren sind auf komplexere Modelle ausgerichtet: Logistische Regression, loglineare Tabellenanalyse, Sterbetafelanalyse, MANOVA-Modelle. Es stellt darüberhinaus eine leistungsfähige Matrix-Programmiersprache zur Verfügung, mit deren Hilfe benutzereigene Prozeduren für den speziellen Bedarf entwickelt werden können.

SPSS Categories: Die statistischen Prozeduren zur Analyse kategorialer Daten sind im einzelnen: Conjoint-Analyse, Korrespondenzanalyse und Skalierungsprozeduren.

SPSS CHAID: Der Name CHAID steht für CHisquared Automatic Interaction Detector und ist eine explorative Technik zur Identifikation von Abhängigkeitsstrukturen in mehrdimensionalen Kontingenztabellen. Das Problem ist die Reduktion einer mehrdimensionalen Kontingenztafel auf eine geringere Zahl von Variablen mit dem Ziel der optimalen Erklärung einer nominalen abhängigen Variablen. Schwerpunkte von CHAID sind erstens die Reduktion der Tabelle durch Überprüfung der Zusammenfaßbarkeit von Kategorien sowie die Auswahl erklärungskräftiger unabhängiger Variablen und zweitens die Identifikation eines geschachtelten Nonstandard-Modells für eine anschließende detailliertere loglineare Analyse. Es ist letztlich eine Technik zur automatischen Identifikation von signifikanten Untergruppen. Die Ergebnisse können sowohl tabellarisch als auch graphisch in Form eines Baumdiagramms dargestellt werden.

SPSS Tables: Das Tabellierungsmodul liefert zwei- und mehrdimensionale Kontingenztabellen in Präsentationsqualität. Die Tools sind flexible Formatie-

rungsmöglichkeiten und unterschiedliche statistische Darstellungsoptionen (Verkettung und Verschachtelung von Variablen innerhalb einer Tabelle, verschiedene Statistiken in Spalten oder Zeilen, frei wählbare Platzierung von Endsummen, Vergabe von Überschriften, Fußnoten und Kommentaren).

SPSS Trends: Dieses Modul umfaßt Verfahren zur Analyse und Prognose von Zeitreihen (ARIMA mit Box-Jenkins Modellen, verschiedene Modelle zur exponentiellen Glättung, Spektralanalyse, multiplikative oder additive Saisonzerlegung für unterjährige, periodische Zeitreihen, Regressionsmodelle mit autokorrelierten Fehlern).

SPSS LISREL: Diese Modul enthält die beiden Prozeduren PRELIS und LISREL, die die Analyse von linearen Strukturgleichungs-Modellen ermöglichen. PRELIS dient der vorbereitenden Datenuntersuchung, wie der Feststellung des Meßniveaus und dem Vorkommen von fehlenden Werten. LISREL ist ein allgemeines Modell zur Untersuchung von linearen Strukturgleichungen in Form der Pfadanalyse. Die Variablen in diesem Gleichungssystem können entweder direkt beobachtete oder nicht beobachtete (sog. latente) Variablen sein. Das Programm ist besonders darauf ausgerichtet, Modelle mit latenten Variablen, mit Meßfehlern in den Indikatoren und mit gegenseitiger Abhängigkeit zu behandeln.

Die einzelnen Prozeduren enthalten eine Vielzahl von Unteroptionen, die den Zugriff auf spezielle Statistiken oder Verfahrensvarianten gestatten.

In Bezug auf die Graphikmöglichkeiten ist SPSS/PC+ völlig inakzeptabel: Integrierte, hochauflösende statistische Graphiken fehlen vollständig. Die Notlösung besteht hier in einer Schnittstelle zu den externen Businessgraphiksystemen CHART oder Harvard Graphics.¹¹ In der aktuellen Version 5.01 existiert zusätzlich eine Schnittstelle zu dem Graphikprogramm AXUM, das auch über einige wissenschaftliche Graphikmöglichkeiten zur Datenanalyse verfügt. Prinzipiell ist eine solche Lösung jedoch nur als Notlösung zu bezeichnen, denn es bedeutet einen zusätzlichen Lernaufwand für ein grundlegend neues Programm, höhere Anschaffungskosten und eine nicht immer gewährleistete vollständige Kompatibilität.

Seit 1993 liegt die Version 5.0 unter Windows vor (die aktuelle Version ist 6.0).¹² Grundsätzlich bietet die Windows Version eines Programms gegenüber der DOS-Version mehrere Vorteile: der augenscheinlichste ist die größere Bedienungsfreundlichkeit durch die graphische Benutzeroberfläche. Ferner ist der Benutzer im Gegensatz zu dem Betriebssystem DOS im Rahmen von Windows nicht auf 640 KB Arbeitsspeicher begrenzt; unter Windows steht (zumindestens theoretisch) der gesamte Arbeitsspeicher für ein Programm zur Verfügung. Au-

¹¹ Es gibt mittlerweile eine eigene Dokumentation über den Transfer zu dem Präsentationsgraphik-PTOgramm Harvard Graphics: SPSS Inc, 1993: Using SPSS/PC+ Graphics with Harvard Graphics, Chicago.

¹² Vgl. die ausführliche Besprechung der Windows-Version von J. Sensch in HSR 19.4 (1994), S. 73-86.

Berdem können Windows-Applikationen viel »enger« zusammenarbeiten: sämtliche Programme orientieren sich an einem einheitlichen Standard, benutzen die gleichen Systemkomponenten (wie z.B. Druckertreiber, Hilfesystem) und können über OLE (»Object Linking and Embedding«) sowie DDE (»Dynamic Data Exchange«) auf einfache Weise Daten austauschen.

Die Dokumentation des Programms ist seit der vor kurzem erschienenen deutschen Windows-Version für das Basismodul auch auf deutsch erhältlich (Übersetzung des Originalhandbuchs) und umfaßt gut 700 Seiten.

Der Programmaufbau ist durch die Windows-Oberfläche noch bedienungsfreundlicher geworden. Anschauliche Symbole, Pull-Down Menüs und leicht verständliche Dialogboxen führen zu einer einfachen Bedienung und sichern eine schnelle Einarbeitung. Wird SPSS für Windows über Dialog-Fenster bedient, werden sämtliche Analyseschritte dokumentiert und in Form der SPSS-eigenen Kommando-Sprache abgespeichert. Einmal ausgeführte Funktionen lassen sich dann wieder jederzeit aufs neue abrufen.

Durch die Nutzung des Windows-Hilfesystems werden die einzelnen Befehle wesentlich ausführlicher erläutert. Die Menüführung und das Hilfesystem sind deutschsprachig. Die Datenanalyse wird im Regelfall unter Benutzung der Windows-Bedienungselemente erfolgen. Da die Batch-Eigenschaften erhalten geblieben sind, ist es auch möglich, Stapeldateien abzuarbeiten. Für umfangreiche Serienuntersuchungen erweist sich die Stapelverarbeitung als wichtige Eigenschaft.

Als Besonderheit ist bei SPSS für Windows der Spreadsheet ähnliche Dateneditor bemerkenswert, den man bei Betätigung der 'Tab'-Taste zeilenweise und bei Betätigung der 'Enter'-Taste spaltenweise durchlaufen kann. Die Eingabe oder die Korrektur von Daten ist damit wesentlich komfortabler geworden.

Darüberhinaus enthält die neue Version 6.0 wichtige neue Features wie die allgemeine Datenbankschnittstelle ODBC (»Open Database Connectivity«), die dem neuen Microsoft-Standard entspricht. Damit ist der Zugriff auf mehr Datenbanken und Anwendungen sichergestellt.

In der um zahlreiche Statistik-Prozeduren erweiterten Programmversion »SPSS für Windows 6.0« (ab Version 5 ist das Programm faktisch die Mainframe-Version SPSS-X mit den entsprechenden Leistungsumfang) sind z.B. differenzierte Verfahren der Ereignisdatenanalyse, die selbst in der Mainframe-Version fehlten, hinzugekommen (Kaplan-Meier-Verfahren, Cox-Regression mit zeitabhängigen Kovariaten).

Erheblich besser als in der DOS-Version sind die Graphikmöglichkeiten. In SPSS für Windows wurde ein eigenständiges Graphikmodul implementiert, mit dem sich zu vielen Prozeduren hochauflösende Graphiken erstellen lassen. Mehr als 50 Diagramme und Graphiken werden auf Wunsch automatisch erstellt. Dazu gehören Balken-, Linien-, Flächen-, Kreisdiagramme, viele Diagramme in 3-D-Darstellung, Histogramme, Boxplots, Punktediagramme. Sie

können auch mit dem leistungsstarken Graphik-Editor bearbeitet werden (z.B. Graphik-Attribute wie Farben, Textgröße, -stil, Kommentare etc.). Das Chart-Karussell ermöglicht sowohl eine Verwaltung als auch ein Betrachten der Graphiken, ähnlich einem Diakarussell in einem Projektor. Über eine Zwischenablage für Text, Zahlen und Graphiken lassen sich die Analyseergebnisse zur Weiterverarbeitung leicht in weitere Windows-Anwendungsprogramme übernehmen.

Die Datenanalyse wird im Regelfall unter Verwendung der Windows-Bedienungselemente erfolgen, aber es ist ebenso möglich, Stapeldateien mit SPSS-Kommandos abzuarbeiten. Wird ein SPSS-Programm in den Editor geladen, so erscheint in der Bedienleiste der 'Run'-Knopf, mit dem das Programm gestartet werden kann.

Die Datenmanagement-Tools sind besonders bemerkenswert. Hier wird nahezu dieselbe Leistungsfähigkeit geliefert wie bei herkömmlichen Datenbankprogrammen. Sicher bieten auch andere Statistik-Programmpakete ähnliche Tools an, besonders benutzerfreundlich sind sie jedoch in der Windows-Version von SPSS implementiert. Ein Beispiel ist das Zusammenfügen mehrerer Dateien. Die Option 'Neue Variablen zusammenfügen' erweist sich als besonders nützlich, um verschiedene Dateien zu einer Auswertung zusammenzufassen, die gemeinsame Variablen enthalten. Diese können als Schlüsselvariablen für die Zusammenfügung von Dateien benutzt werden.

Die Dokumentation umfaßt zwei Bände: SPSS for Windows. Base System User's Guide (828 Seiten); SPSS Base System. Syntax Reference Guide (940 Seiten). Jedes Erweiterungsmodul wird in einer gesonderten Dokumentation beschrieben. Festzuhalten ist: SPSS hat den Sprung in die Windows-Welt mit einer überzeugenden Portierung von SPSS-X vollzogen.

Preisklasse: ca. DM 2500,- (Basispaket SPSS für Windows); jedes weitere Modul kostet DM 1100,-

Vertrieb: SPSS GmbH Software, Rosenheimer Straße 30, 81669 München.

BMDP

Das zweite hier vorzustellende, insbesondere an vielen Hochschulrechenzentren gut bewährte und häufig verwendete Statistik-Programmpaket ist BMDP: »BioMedical Computer Programs P-Series«. BMDP ist eine Fortentwicklung des wohl ältesten Statistik-Programmsystems BMD (»BioMedical Computer Program«), das bereits 1963 eingeführt wurde. Der Name verweist auf den ursprünglich in der Medizin und der Biologie angesiedelten Verwendungsbereich. Heute wird dieses Programm jedoch in nahezu allen Erfahrungswissenschaften eingesetzt. Im Vergleich zu den Konkurrenzprodukten unterscheidet sich BMDP hinsichtlich seines Aufbaus: Es ist kein Programmsystem, sondern ein Programmpaket, in dem 44 verschiedene, aber in gleicher Syntax zu

benutzende statistische Programm-Module auswählbar sind. BMDP ist für verschiedene Rechnertypen und Betriebssysteme verfügbar. Eine menüorientierte PC-Version (BMDP/PC Classic System 7.0) in völlig überarbeiteter Aufmachung liegt seit 1990 vor. Eine direkte menügesteuerte Erstellung von Programmbefehlen, wie es andere Pakete mit prozeduralen Kommandosprachen anbieten, ist jedoch nicht verwirklicht. Der durch die Oberfläche zur Verfügung gestellte Benutzersupport muß daher eher als spärlich bezeichnet werden. Die Dokumentation umfaßt drei Bände:

Vol. 1: BMDP Statistical Software Manual, Vol. 2: BMDP Statistical Software Manual und Vol 3: BMDP Data Manager.

BMDP kann nach den individuellen Erfordernissen des Anwenders modular zusammengestellt werden, indem aus den 44 statistischen Prozeduren lediglich eine Auswahl einzelner Programme gekauft oder gemietet wird. Seit 1993 ist auch eine Windows-Version auf dem Markt (BMDP New System für Windows).

BMDP stellt ein breites Angebot an statistischen Verfahren zur Verfügung. Die Anwendungsgebiete reichen weit über den ursprünglich naturwissenschaftlichen Bereich hinaus. Aus dem statistischen Methodenspektrum seien hier exemplarisch erwähnt

- Elementare Auswertungen (Häufigkeitstabellen, Histogramme, Korrelationen, Kontingenztabellen, einfache Varianzanalyse, Mittelwertvergleiche in zwei Gruppen);
- mehrfache Varianz- und Kovarianzanalyse (mit oder ohne Meßwiederholungen), multivariate Varianz- und Kovarianzanalyse;
- hierarchische Clusteranalyse (von Variablen oder von Untersuchungseinheiten);
- Faktorenanalyse (einschließlich der Booleschen Faktorenanalyse für dichotome Variablen), Diskriminanzanalyse (einschließlich der schrittweisen), kanonische Korrelation;
- Beschreibung und Schätzung von fehlenden Werten;
- einfache und multiple Regressionsanalyse, schrittweise Regressionsanalyse, nichtlineare Regression, Ridge-Regression, multiple Regression mit Hauptkomponenten, Polynom-Regression, multivariate Regression;
- mehrdimensionale Kontingenztabellenanalyse (loglineare Modelle);
- Analyse diskreter abhängiger Variablen (logistische Regression mit dichotomen oder polytomen abhängigen Variablen; logistische Regression mit ordinalen abhängigen Variablen, schrittweise logistische Regression);
- Ereignisdatenanalyse (Sterbetabellen, Überlebensfunktionen, Überlebensfunktionen mit Kovariaten (Cox-Regression), Überlebensfunktionen mit zeitabhängigen Kovariaten);
- nichtparametrische Tests;
- Zeitreihenanalyse (uni- und bivariate Spektralanalyse, Box-Jenkins-Zeitreihenanalyse);

- Datenmanagement, Verarbeitung von nichtrechteckigen Datenmatrizen.

Zusätzlich öffnet BMDP/PC-90 - die jüngste menueorientierte PC-Version - mit dem Modul »Data Entry« einen breiten Weg zur EDV-gestützten Datenerfassung und -Verwaltung.

Die Version BMDP/386 mit »Extended Memory« für die Nutzung von 4 MB oder 8 MB Hauptspeicher schränkt die Notwendigkeit, Großrechner benutzen zu müssen, auf jene wenigen Fälle ein, bei denen eine extrem große Anzahl von Untersuchungseinheiten zu analysieren sind. Für PCs mit dem Betriebssystem UNIX steht 'BMDP for X Windows' zur Verfügung. Die Statistik-Programme und die Syntax der BMDP-Anweisungen sind für alle Versionen gleich.

Das schon erwähnte *New System* ist eine Neuentwicklung für Windows und hat nichts mit dem 'normalen' BMDP (dem *Classic System*) zu tun. Das Methodenspektrum ist noch vergleichsweise schmal. Es umfaßt neben den üblichen elementaren Statistiken z. Zt. nur die Analyse zweidimensionaler Kontingenztabellen, lineare Regression, t- und U-Tests sowie die ein- und zweifaktorielle Varianzanalyse. Die wesentliche Neuerung besteht darin, das BMDP New System als sogenanntes 'Frontend' für das 'normale' BMDP ('Classical System') benutzt werden kann, d.h. man kann unter Windows und dort mit dem New System arbeiten und dabei aus dem Programm heraus das gewohnte BMDP-Programm verwenden, ohne z.B. das New System verlassen zu müssen oder sich um die Datenübergabe zwischen den beiden Systemen kümmern zu müssen.

Preisklasse: ca. DM 3300,-.

Vertrieb: Statistical Software Ltd., Cork Technology Park, Model Farm Road, Cork, Ireland.

CSS: STATISTICA

Das Programmsystem CSS: STATISTICA ist die Weiterentwicklung des seit 1987 auf dem Markt befindlichen Programms CSS 2 der Firma StatSoft/USA. CSS ist die Abkürzung von »Complete Statistical System«. Das Datenanalysesystem ist speziell für den PC entwickelt worden. Es wird unter STATISTICA/W für das Betriebssystem Windows, STATISTICA/DOS für DOS und unter STATISTICA/MAC für Macintosh angeboten. Das komplette Programmpaket benötigt 14 MB Plattenplatz. Zudem wird eine Sparversion angeboten (»Quick STATISTICA«), die neben allen graphischen und Datenmanagementfunktionen nur Basisstatistik-Prozeduren aufweist. Als Dokumentation wird ein 150 Seiten umfassendes Handbuch »Quick Reference« sowie drei insgesamt 2506 Seiten umfassende Bände »Conventions & Statistics I«, »Graphics« und »Statistics II« mitgeliefert.

Mit CSS: STATISTICA können umfangreiche Datenbestände verwaltet und statistisch analysiert werden. Ein besonderes Kennzeichen des Datenanaly-

seprogramms sind die vielfältigen Möglichkeiten der graphischen Datenanalyse. Das Programm ist universell ausgerichtet und unterstützt Anwendungen in allen Bereichen von Forschung und Lehre sowie Verwaltung und Wirtschaft.

STATISTICA arbeitet die einzelnen Operationen mit einer hohen Geschwindigkeit ab und liefert sehr präzise Ergebnisse. Die Zahl der zu bearbeitenden Untersuchungseinheiten ist praktisch unbegrenzt. Sie wird lediglich von der Speicherkapazität des PC beschränkt.

Die einzelnen Menüpunkte und Dialogboxen sind intuitiv verständlich. Es steht auch eine Online-Dokumentation zur Verfügung, über die Hilfstexte abgerufen werden können. Sie geben Auskunft über die einzelnen Programmteile und beschreiben zum Teil sehr ausführlich die entsprechenden Verfahren. Alternativ zu den Menues verfügt das Programm über eine Kommandosprache SCL (»STATISTICS Command Language«), die von der Logik den Kommandosprachen von z.B. SPSS oder BMDP entspricht. In bezug auf den Austausch mit anderen Programmen zeigt sich STATISTICA sehr flexibel: es kann Excel-, Lotus-, QuattroPro-, Paradox-, SPSS-Portable, ASCII-, DIF-Formate sowie im sog. Megafile-Manager-Format importieren.

Innerhalb des Programms gibt es verschiedene Datei-Formate, die jeweils in unterschiedlichen Fenstern dargestellt werden können: Spreadsheet-Dateien enthalten die Daten, Scrollsheet-Dateien enthalten die Ergebnisse der statistischen Analyse. Graphiken werden in einem dritten Format gespeichert und schließlich kann man die Ergebnisse einer kompletten Sitzung (oder eines Teils) noch in einer Textdatei mitprotokollieren lassen.

Das Programm besteht aus Runtime-Modulen, welche die verschiedensten Bereiche abdecken. Startet man das Programm ohne nähere Angabe, erscheint ein sog. Modulswitcher, von dem aus ein bestimmtes Auswertungs-Modell (d.h. eine Statistik-Prozedur) aufgerufen werden kann. Die Programm-Module umfassen sowohl einfache deskriptive als auch komplexe multivariate statistische Auswertungen von Daten und ermöglicht vielseitige graphische Darstellungen. Das Methodenspektrum umfaßt auch zahlreiche Modelle zur Analyse von Zeitreihen. Dabei sind innerhalb der einzelnen Verfahrenskomplexe neben den bekanntesten Verfahren oft auch spezielle Verfahrensvarianten, wie etwa im Bereich der Varianzanalyse in Form einer generalisierten MANOVA-Prozedur zu finden. Das Programm verfügt über eine durchgängig menügesteuerte Benutzeroberfläche, die sehr übersichtlich ist.

Neben dem umfassenden Methodenangebot zeichnet sich CSS vor allem auch durch seine voll integrierten, mächtigen Graphikmöglichkeiten aus: Jedes Ergebnis einer statistischen Analyse ist direkt aus dem jeweiligen Statistikmodul heraus in eine graphische Darstellung umsetzbar, die dann sofort auf vielfältige Weise den eigenen Ansprüchen und Vorstellungen entsprechend gestaltet werden kann. Es werden jedoch nur die bekannten Business-Graphiken hochauflösend erstellt Box-Plots, Histogramme oder Streuungsdiagramme sind dagegen nur als Printer Plots implementiert.

Eine weitere Stärke von CSS stellt die an die Pascal-Syntax angelehnte, interne Programmiersprache dar. Sie erlaubt auf der Grundlage der in CSS eingegebenen Daten eine eigene Entwicklung nicht explizit enthaltener Modelle oder Modellvarianten. Neben der Berechnung ist auch die Programmierung der Ausgabeergebnisse analog zu Pascal möglich.

Inzwischen wird STATISTICA auch in einer speziellen Windows-Version angeboten (Release 4.3A). Nach dem Start bieten Icons eine Vielzahl von einfach zu handhabenden Bedienelementen an. Dabei werden alle Register einer graphischen Benutzeroberfläche ausgeschöpft, um den Benutzer eine interaktive Arbeit zu ermöglichen. Sämtliche Programmoptionen sind per Mausklick einstellbar. Die Ergebnisse werden als Icons im Programmfenster abgelegt und stehen nach dem Verlassen des Programms beim Neuaufruf wieder zur Verfügung.

Nach wie vor stehen sowohl die Kommandosprache als auch die Programmiersprache der DOS-Version zur Verfügung. Damit lassen sich auch unter Windows sehr komplizierte Auswertungsalgorithmen programmieren. Business-Graphiken werden in Fenster abgelegt, die sich zu neuen Fenstern zusammenfassen lassen. Sie können dann beliebig graphisch aufgearbeitet werden.

Preisklasse: ca. DM 2300,-.

Vertrieb: Loll + Nielsen, Software Vertrieb, Quickbornstr. 3, 20253 Hamburg.

ALMO

Der Name ALMO leitet sich als Abkürzung aus dem allgemeinen linearen Modell ab, ein auf Variablen beliebigen Meßniveaus verallgemeinertes Regressionsverfahren. Das allgemeine lineare Modell umfaßt als Submodelle die klassischen Verfahren der Regressionsanalyse, der Varianz- und Kovarianzanalyse, der Diskriminanzanalyse etc., um nur einige zu nennen.

ALMO wurde ursprünglich auf dem Großrechner entwickelt und dann in den Mikrocomputerbereich portiert. Dabei wurde ein Programmsystem geschaffen, welches von Atari ST über OS/2 nun auch für Windows 3.1 erhältlich ist. Seit Oktober 1988 wird es auch für Mikrocomputer angeboten. Das von Prof. Dr. Kurt Holm an der Universität Linz entworfene Programm ist heute ein hoch integriertes, modular konzipiertes, aus einer Vielzahl von Unterprogrammen bestehendes System.

Die Programmsteuerung ist in der Windows-Version konsequent an die graphische Benutzeroberfläche angepaßt. Es gibt fertige Programm-Masken, die der Benutzer laden und nach seinen spezifischen Anforderungen anpassen kann. Solange der Benutzer die kommandoorientierte ALMO-Syntax noch nicht beherrscht, schreibt er in diese Masken nur einige Eingabewerte entsprechend seinen Auswertungszielen und startet dann das Programm. Mitten im

Programmtext sind anklickbare »buttons« angebracht, über die jederzeit Hilfen aufgerufen werden können. Diese Sonderform einer kontextsensitiven Hilfe ist praktisch ein Bestandteil des ALMO-Programmeditors.

Neben einer Vielzahl von statistischen Prozeduren, die z.B. auch für den Bereich der Zeitreihenanalyse zahlreiche Optionen bieten, enthält das Programm über 150 Beispielprogramme, die zum großen Teil Statistik-Lehrbüchern entnommen sind. Die Beispiele sind ausführlich kommentiert, so daß dem Benutzer sehr schnell der formal-statistische Kern der einzelnen Prozeduren transparent wird. Mit einem Mausklick kann er die Beispielprogramme starten und rechnen. Die Programmdokumentation ist sehr umfangreich (etwa 1000 Seiten DIN-A4) und nimmt an vielen Stellen Bezug auf die Algorithmen der jeweiligen Analyseprozedur.

Das Programmsystem enthält darüberhinaus eine einfache Programmiersprache (in deutscher Syntax), mit der die verschiedenen Verfahren aufgerufen und Variablenwerte umkodiert werden können. Mit Hilfe dieser Sprache kann der Benutzer auch komplexe Programme, wie z.B. Matrix-Operationen, programmieren.

ALMO umfaßt ferner eine vollständige, programmierbare relationale Datenbank, die zahlreiche Funktionen enthält (z.B. Anhängen von Datensätzen an bestehende Dateien, Sortieren von Dateien, relationale Datenbank-Operationen etc.).

Das Programmsystem bietet unterschiedliche Optionen im Bereich der Business-Graphik, enthält darüberhinaus auch graphische Ergebnisdarstellungen für spezielle Auswertungsprozeduren (z.B. der Faktorenanalyse, der Clusteranalyse und der Korrespondenzanalyse). Unter Benutzung von 'Schiebern' können 3-dimensionale Graphiken im Raum gedreht werden.

Die Ergebnisse einer Auswertung können über das Clipboard in alle Windows-Anwendungen exportiert oder über einem Drucker ausgegeben werden.

Preisklasse: ca. DM 500,-.

Vertrieb: Prof. Dr. Kurt Holm, Universität Linz, Österreich; Priv.: Am Schlöblberg 8, A-4060 Leonding.

NSDstat+

Das Statistikprogramm NSDstat+ ist ein speziell für PCs entwickeltes Analysepaket, das von dem Norwegian Social Science Data Service in Bergen entwickelt wurde und im deutschsprachigen Raum über ZUMA (Mannheim) zu beziehen ist. Die deutsche Version dieses Programms (einschließlich Handbuch) wurde im Rahmen eines gemeinsamen GESIS-Projektes unter der Federführung von ZUMA in der Außenstelle Berlin erstellt. Mit NSDstat+ können selbst große Datensätze extrem schnell ausgezählt, tabelliert und die Er-

gebnisse graphisch dargestellt werden. An Geschwindigkeit und Einfachheit der Handhabung übertrifft es die großen Standardstatistikpakete um ein Vielfaches. Daher ist NSDstat+ ein besonders geeignetes Instrument für die Datenexploration und deskriptive Standardanalyse. Darüber hinaus ist es wegen seines sehr guten Benutzerkomforts auch für Statistikurse geeignet

Eine besondere Eigenschaft von NSDstat+ ist die enge Verknüpfung aller numerischen Ausgaben (Tabellen) mit graphischen Darstellungen, die direkt und ohne langwierigen Umweg über externe Graphikprogramme erzeugt werden können. Die Graphiken können auf Druckern ausgegeben werden. Es ist auch möglich, thematische Karten für Datensätze zusammenzustellen.

Das Programm enthält die wichtigsten statistischen Prozeduren, die für eine ausführliche Datenexploration neuer Datensätze völlig ausreichend ist. Viele Forschungsfragen lassen sich bereits mit dem gelieferten Analysespektrum beantworten. Zu jeder statistischen Prozedur gibt es Informationen in Hilfefenstern zu den statistischen Eigenschaften bzw. zu den verwendeten Formeln. Weiterhin sehr hilfreich sind die Möglichkeiten der statistischen Deskription (z.B. das Einblenden einer Normalverteilung in die Graphik einer empirisch ermittelten Häufigkeitsverteilung).

Preisklasse: ca. DM 450,-.

Vertrieb: Zentrum für Umfragen, Methoden und Analysen (ZUMA), B2, 1, 68072 Mannheim.

P - STAT

Das Programm P-STAT ist ein dialogorientiertes statistisches Informationsanalyse- und Datenbanksystem für Computer verschiedenster Größenordnungen, angefangen beim Mikrocomputer bis hin zum Großrechner. Im Grundkonzept baut es ebenfalls auf einer prozeduralen Kommandosprache auf, die auf den verschiedensten Rechnersystemen einheitlich ist. Auf Großrechenanlagen wird P-STAT seit mehr als 20 Jahren im wissenschaftlichen sowie kommerziellen und administrativen Bereich eingesetzt.

Nach dem Programmstart steht dem Benutzer zunächst ein reines kommandozeilen-orientiertes System zur Verfügung. Einsteigern empfiehlt sich die Einbindung des optionalen Menüsystems, welches zwar keinen hohen Standard hat, aber zumindest einen DOS-Minimalkomfort bietet. Mit dem (etwas antiquiert anmutenden) Menüsystem lassen sich die Kommandozeilen für die Befehlsdatei interaktiv zusammenstellen. Dies ist zumindest einfacher als die Verwendung des ASCII-Editors. Mit dem Menüsystem steht ein syntaxorientierter Editor für Kommandodateien zur Verfügung, der die fehlerfreie Erzeugung auch komplizierter Befehlsdateien ermöglicht.

Bei der auf Stapelverarbeitung ausgerichteten Philosophie des Programms kommt es nicht unerwartet, daß interaktive Graphikmöglichkeiten fehlen. Nicht

vorgesehen sind integrierte hochauflösende Graphiken. Es besteht lediglich eine Schnittstelle zu CHART von Microsoft.

Die von P-STAT abgedeckten Bereiche statistischer Auswertungen umfassen die beschreibende und schließende Statistik, Tabellierungsverfahren, semi-graphische Darstellungen, explorative Datenanalyse, multivariate Verfahren und Methoden der Zeitreihenanalyse und Prognoserechnung. Alle diese Prozeduren können durch eine Vielzahl von Optionen gesteuert und den Benutzeranforderungen entsprechend angeglichen werden. Der statistische Leistungsumfang entspricht in etwa dem der anderen großen klassischen Statistiksysteme.

Das P-STAT-Datenbankmanagementsystem (PDMS) besitzt eine eigene, sehr mächtige Programmiersprache (PPL), die dem Benutzer eine Lösung des Daten- und Dateimanagements erlaubt. Es ist darüberhinaus möglich, die Zusammenfassung von PPL-Sequenzen in Makros einzubinden und in einer Bibliothek abzulegen.

Neben dem Statistik- und Datenbanksystem enthält P-STAT schließlich noch einen eigenen Texteditor, der es ermöglicht, Programm- und Datendateien zu bearbeiten.

Preisklasse: ca. DM 3000,-.

Vertrieb: VGSPS mbH, Vertriebsgesellschaft für statistische Programmsysteme, Pützchens Chaussee 60, 53227 Bonn (Beul).

SYSTAT mit SYGRAPH

Das Programm SYSTAT ist ein universelles Programmsystem zur statistischen Datenanalyse. Es wurde vor über 20 Jahren zunächst als Großrechner-Programm entwickelt und seitdem bis zur Version 5 für Mikrocomputer weiterentwickelt. Das System weist eine perfekt auf den PC (bzw. Macintosh) zugeschnittene Benutzeroberfläche auf. Die Dokumentation besteht aus einem 255 Seiten umfassenden Handbuch »Getting Started«, das in die grundlegenden Prinzipien des Programms einführt, einem 248 Seiten umfassenden Band »DATA«, in dem das Daten-Management und die Datentransformations-Optionen beschrieben werden, und einem 677 Seiten umfassenden Handbuch »Statistics«. Inzwischen existiert auch eine Windows-Version, die ebenso wie SPSS für Windows noch bedienungsfreundlicher ist.

In der PC-Version wird der Anwender mit Hilfe von Pull-Down-Menüs vom allgemeinen Anwählen eines Auswertungsverfahrens bis hin zu der Menge aller verfügbaren Unteroptionen geführt. Dem Anwender wird freigestellt, ob er mit der Menüsteuerung arbeiten will oder ob er die Zeilensteuerung bzw. die Programm-Steuerung mit mehreren Programm-Zeilen bevorzugt (direkte Eingabe einer Befehlsfolge und Starten als Batch-Job). Das Menü-System erlaubt eine Abspeicherung der angewählten Befehle in eine Kommandodatei, bei der

ein Steuerprogramm angelegt und abgespeichert wird. Das Datenmanagement und der Daten-Austausch mit anderen Programmen entsprechen weitgehend denen von SPSS/PC+. Interaktiv können Daten in einem Spreadsheet-Modul »Edit« bearbeitet werden (analog zu SPSS für Windows).

SYSTAT ist ein Programm-System, das aus einer Vielzahl von Einzel-Modulen besteht, von denen jedes für sich allein betriebsfähig ist. Das System enthält nahezu alle wichtigen statistischen Verfahren, sei es im Bereich der deskriptiven, der schließenden oder der multivariaten Statistik. Hinsichtlich der Konnexität der in SYSTAT enthaltenen Statistik-Prozeduren ist es sicherlich eines der vollständigsten und anspruchvollsten Statistik-Programmpakete. Erweiterungs-Module zu speziellen Themengebieten erhöhen den Leistungsumfang und runden das Paket ab. Um nur einige zu nennen: MESOSAUR analysiert Zeitreihendaten und beinhaltet hochentwickelte interaktive Graphiken für diese Aufgabe; SURVIVAL umfaßt ein Verfahrensspektrum zur Ereignisdatenanalyse; LOGIT umfaßt die Klasse der Analyse diskreter abhängiger Merkmale (logistische Regressionsanalyse etc.); SAMPLE dient der Stichprobenbildung. Das Programm-Modul SYTRAN (Version 3.0) bietet eine Schnittstelle zu einer breiten Anzahl von Programmen (Kovertierung von Files für verschiedenste Datenbankprogramme und Statistik-Programmpakete). Bei der Umsetzung bleiben die Missing Value-Codes erhalten, die Labels für Variablen und Werteausprägungen werden jedoch nicht übernommen. Die Erweiterungsmodule sind Runtime-Module, d.h. sie können unabhängig von dem Basisprogramm eingesetzt werden (z.B. SYTRAN für File-Konvertierungen).

Eine weitere hervorzuhebende Stärke von SYSTAT hegt in seinen in SYGRAPH zusammengefaßten Graphik-Programmen. In der Version 5 lassen sich hochauflösende ein- bis dreidimensionale Graphiken erstellen. Während von der Vielfalt der Graphiken einige der bereits besprochenen Pakete durchaus mithalten können, ist SYGRAPH sowohl im Hinblick auf die Qualität der Graphiken als auch auf deren Kombinierbarkeit kaum zu schlagen. Wer die ganze Leistungsbreite voll ausschöpfen will, muß hier jedoch von der Menüoberfläche Abschied nehmen und auf die Kommandoebene zurückgreifen. Dann erhält der Anwender allerdings auch ein hochwertiges Analyse- und Präsentationswerkzeug.

Die Programmierfähigkeit und die hervorragenden Graphikfähigkeiten sind in der Windows-Version (V. 5.02) vollständig portiert worden. Schon das DOS-Programm mit seiner SAA-ähnlichen Menüstruktur war recht benutzerfreundlich. Die Windows-Fassung bietet darüberhinaus noch eine Optimierung der Programmsteuerung durch den Gebrauch aller Möglichkeiten der Windows-Oberfläche. Das Ergebnis ist eine sehr funktionale, einfach zu überblickende Benutzeroberfläche, die extensiv mit Icons arbeitet. Unterhalb der Menüleiste befindet sich eine Knopfleiste mit Icons, die als 'Quick'-Tasten bezeichnet werden. Die Erste Analysen sind durch ein oder zwei Mausklicks auf die Icons möglich. Für detaillierte Analysen ist dennoch der Weg über die Menüleiste

empfehlenswert, wobei auch hier Icons den richtigen Weg weisen. Das Programm arbeitet mit insgesamt fünf Fenstern. Die Daten gibt man im 'Worksheet'-Fenster ein, kann sie hier verändern und ansehen, während Kommandodateien und statistische Ergebnisse im 'Notepad'-Fenster eingegeben, ediert und gespeichert werden. Die Ergebnisse statistischer Analysen werden im 'Main'-Fenster ausgegeben. Das 'Graph'-Fenster zeigt schließlich die gewünschten Graphiken an. Zusätzlich steht noch das Clipboard für Kopierzwecke zur Verfügung (als Schnittstelle für den Transfer von Analyseergebnissen zu anderen Windows-Programmen).

Neben der Datenübertragung über das Clipboard bietet SYSTAT die bidirektionale Umwandlung zwischen den Formaten Lotus 123, Lotus Symphony, dBase oder Portable (einfaches **ASCn**-Format). Dazu kommen die Windows-Graphikformate Bitmap und Metafile.

Es sei jedoch angemerkt, daß die Windows-Version über keine Schnittstelle zu den DOS-Erweiterungsmodulen (z.B. PROBIT, LOGIT, SURVIVAL, MESOS AUR etc.) verfügt. Der Benutzer muß vor dem Start eines Zusatzmoduls die Windows-Version verlassen und startet dann das entsprechende DOS-Spezialmodul für seine Datenauswertung.

Im September 1994 hat SPSS Inc. die Firma SYSTAT Inc. übernommen. SYSTAT-Produkte werden daher seit Anfang 1995 von SPSS Inc. vertrieben. Aufgrund einer stärker gewordenen Aufteilung innerhalb des wissenschaftlichen Sektors, erwartet SPSS eine größere Hebelkraft durch cross-selling bestimmter Module. SPSS plant, die Entwicklung und Vermarktung beider Produkte fortzusetzen, wobei jedes für sich ständig überarbeitet wird. Geplant ist die Verwendung von gemeinsamen internen 'application programming interfaces' (APIs). Die Verwendung der gleichen Standard-Interfaces in beiden Produkten ermöglicht es, neue Features, bzw. einige existierende Features, gemeinsam in beiden Produkten zu nutzen. Die neueste Version von SYSTAT für Windows wird Ende 1995 erscheinen. Eine Besprechung erfolgt demnächst in der HSR.

Preisklasse: ca. DM 2000,-.

Vertrieb: SPSS GmbH Software, Rosenheimer Str. 30, 81669 München.

STATGRAPHICS

STATGRAPHICS, das derzeit in Version 5.1 vorliegt, ist ein vielseitiges Statistik-Programmsystem für den PC mit integrierten Graphikfunktionen hoher Auflösung. Im Gegensatz zu anderen Statistik-Programmsystemen ist STATGRAPHICS nicht lediglich der PC-Abkömmling eines älteren, batchorientierten Mainframe-Softwarepakets, sondern ein relativ kompaktes, modular aufgebautes Programmsystem, das speziell auf die Möglichkeiten des PCs zugeschnitten ist und somit die Wünsche des PC-Benutzers nach einfacher Bedie-

nung erfüllt. Die Benutzerführung erfolgt mit Hilfe eines Menüsystems, das den Anwender durch alle verfügbaren Prozeduren führt. Die Bedienung erfolgt ausschließlich über die Tastatur (d.h. keine Menüleiste, keine Mausunterstützung). Man wählt mit den Cursortasten einen Menüeintrag aus und bestätigt mit der 'Enter'-Taste. Dann taucht das nächste Untermenü auf. In den einfach zu aktivierenden Eingabemenüs sind so lediglich die gut kommentierten Felder auszufüllen oder eine Option zu wählen.

Neben einer Vielzahl von statistischen Prozeduren, die z.B. in dem Bereich der Zeitreihenanalyse zahlreiche Optionen bieten, ermöglicht dieses Programm auch die Anwendung einiger mathematischer Verfahren wie numerische Integration und Differentiation beliebiger Funktionen, Nullstellenbestimmung und Lösung linearer Gleichungssysteme. Darüber hinaus kann der Anwender mit der Programmiersprache APL weitere Verfahren installieren und bereits bestehende Statistikmodule erweitern. Das statistische Auswertungsspektrum umfaßt:

- deskriptive Statistik (Häufigkeitstabellen, Häufigkeitshistogramme, Verteilungskurven);
- explorative Statistik (Box- und Whisker-Plots, multiple Box- und Whisker-Plots, suspended Rootogram, hängende Balkenhistogramme);
- statistische Prüfverfahren (Ein-Gruppen-Verfahren, Zwei-Gruppen-Verfahren, Zwei-Gruppen-Vergleich, Anpassen von Verteilungen);
- Varianzanalysen (ein- und mehrfaktorielle Varianzanalyse, Analyse von Nested Designs);
- Regressionsanalyse (einfache und multiple Regression, multivariate Regression);
- mehrdimensionale Tabellenanalyse (loglineare Modelle);
- Korrelationsanalyse (Rangkorrelation, Korrelation, partielle Korrelation, kanonische Korrelation, Faktorenanalyse);
- Zeitreihenanalyse (Brown's Exponentialglättung, Holt's lineare Exponentialglättung, Trendanalyse; gleitender Mittelwert, polynomiale Glättung, offene und geschlossene QSplines).

Der interaktive Teil von STATGRAPHICS läßt dem Anwender die Wahl unter mehr als 50 verschiedenen graphischen Darstellungsarten: herkömmliche Torten- und Balkendiagramme, Histogramme, 3D-Liniendarstellungen, Topogramme etc. Alle Darstellungsarten sind integraler Bestandteil der Bedieneroberfläche. Dadurch können die verarbeiteten Daten unmittelbar dargestellt und die Darstellungen jederzeit modifiziert werden.

Seit Januar 1992 wird neben der Version 5.1 auch STATGRAPHICS/5.2 (STATGRAPHICS PLUS) speziell für PCs mit den Intel-Prozessoren 386/486 angeboten. Die PLUS-Version ist gegenüber STATGRAPHICS 5.1 nicht nur wesentlich schneller, sondern läßt auch die Verarbeitung weit größerer Datenmengen zu. Die Systemarchitektur, die ein Halten der gesamten Auswertungs-

daten im Speicher vorsieht, schränkte den Leistungsumfang des Systems im Hinblick auf die Auswertung größerer Datenmengen erheblich ein.

Die aktuellste Version 7, die derzeit nur in englischer Fassung erhältlich ist, hat der Hersteller mit einer benutzerfreundlicheren Oberfläche versehen, darüberhinaus wurde auch der Funktionsumfang erhöht. Seit Anfang 1994 ist auch eine Windows-Version auf dem Markt. Im weiteren Verlauf des Jahres 1995 ist geplant, dem Basis-System (deskriptive Statistik, multiple Regression, ANOVA etc.) dann u.a. Module zur Zeitreihenanalyse und zu weiterführenden multivariaten Datenanalyseverfahren folgen zu lassen.

Preisklasse: ca. DM 2600,-.

Vertrieb: STSC, Inc., 2115 East Jefferson Street, Rockville, Maryland 20852, USA.

Repräsentant für Deutschland: ISE DATA, Herrrainweg 5, 63067 Offenbach/M.

SAS

Das Programm SAS (»Statistical Analysis System«) ist weit mehr als nur ein Statistik-Programmsystem. SAS bezeichnet seine Software als 'Information Delivery System'. Das SAS System ist ein modular aufgebautes, integriertes Anwendungsentwicklungssystem, das eine vollständige Kontrolle der zu verarbeitenden Daten ermöglicht. Die SAS Sprache (Sprache der 4. Generation) in Verbindung mit fertigen Auswertungsprozeduren bietet ein strategisches Werkzeug zur Informationsverarbeitung. Grundlage des SAS Systems bilden die vier Funktionen, die eine Vielzahl von Anwendungsmöglichkeiten bieten:

Datenzugriff	(transparent auf gängige Datenbanken und diverse Rohdatenformate)
Datenverwaltung	(Datenmanagementfunktionen)
Datenanalyse	(deskriptive, multivariate Statistik)
Präsentation	(Tabellen, Berichte, Graphiken etc.)

SAS wurde von vornherein als ein statistisches Analysesystem zum Einsatz in unterschiedlichen Disziplinen entwickelt. So wird es heute im wirtschafts- und sozialwissenschaftlichen Bereich, in den Naturwissenschaften sowie in der Wirtschaftspraxis, z.B. bei Großbanken, Industrieunternehmen usw. eingesetzt. SAS läuft unter vielen Großrechnerbetriebssystemen, auf Workstations unter UNIX sowie auf PCs in der Version 6.04 unter dem Betriebssystem MS-DOS und in der Version 6.08 unter OS/2 und Windows. Es wird in verschiedenen Modulen angeboten, die alle auf das sog. Basis-Modul zurückgreifen. Das Basis-Modul dient der Speicherung und der Modifikation von Daten und bietet darüberhinaus deskriptive Statistiken und einfache graphische Darstellungsmöglichkeiten. Weitere angebotene Module sind z.B. das SAS/STAT-Modul,

das wichtige Bereiche der statistischen Methoden abdeckt. Darüberhinaus gibt es noch eine ganze Reihe weiterer Module, die mit jeder Releaseänderung laufend gewartet und erweitert werden, wie z.B.

- SAS/AF zum interaktiven Einsatz der SAS-Anwendungen,
- SAS/FSP, ein interaktiver, menuegesteuerter Full-Screen-Editor zur Dateineingabe und Datenmodifikation, zur Ausführung von DAS-Files und zur integrierten Textverarbeitung,
- SAS/GRAPH, ein Modul zur Erstellung von Präsentationsgraphiken,
- SAS/IML als interaktive, matrixorientierte fortgeschrittene Programmiersprache, um u.a. mathematische Operationen sowie komplexe statistische Modelle schnell in eigene Programme umzusetzen.

SAS arbeitet seit längerem mit einer selbstentwickelten Fenstertechnik. Die interaktive Menüoberfläche ermöglicht einen schnellen Einstieg. Die Benutzeroberfläche der PC-Version orientiert sich dabei ebenso wie die Funktionsstabenbelegung an der Großrechnerversion für verschiedene IBM-Betriebssysteme. Es stehen drei Fenster zur Verfügung, wovon eines den Editor zur Eingabe der Kommandosequenzen, das zweite die LOG-Datei und das dritte den Editor mit den generierten Ergebnissen enthält. Die Erstellung einer Auswertung erfolgt analog zu allen anderen vom Großrechner kommenden Systemen über eine prozedurale Kommandosprache. Die PC-Version unterstützt dabei den Anwender mit einem Zusatzpaket ASSIST, die eine über Fenster gesteuerte Zusammenstellung der Befehlssequenzen erlaubt. Sie ermöglicht einen schnellen Einstieg in das System und unterstützt das Erlernen der SAS-Kommandosprache.

Bezüglich der verfügbaren statistischen Verfahren ist SAS sowohl in der Breite der angebotenen Verfahren als auch in der Vielfalt der Optionen in den einzelnen Unterprogrammen das zur Zeit umfassendste System. Eine große Zahl von Zusatzprozeduren, die in der SAS-User-Group entwickelt werden, ergänzt laufend das Gesamtsystem und sorgt damit für maximale Aktualität des Methodenspektrums. Innerhalb der einzelnen Module können verschiedene Programmroutinen, die sog. PROCs, aufgerufen werden, wobei eine Vielzahl von Optionen durch den Nutzer definiert werden können.

In dem Zusatzpaket SAS/GRAPH sind hochauflösende Graphiken implementiert, die von exzellenter Qualität sind. Der Schwerpunkt liegt dabei in dem Bereich der Businessgraphiken. Die Graphiken lassen sich beliebig vergrößern oder verkleinern.

SAS ist ein gigantisches Programmsystem, welches als reine Statistik-Software einer völligen Fehlbewertung unterliegen würde. Der Umfang dieses Systems wird auch an der Speicherplatzbelegung deutlich: über 100 Megabyte für das Gesamtsystem.

Von den meisten anderen am Markt angebotenen Statistikpaketen unterscheidet sich SAS dadurch, daß von den Anwendern kein Eigentum an der

Software erworben werden kann. Es handelt sich vielmehr um eine zeitlich beschränkte Nutzerlizenz, die z.B. für ein Jahr erworben werden kann. Dadurch fallen periodische Kosten für dieses Softwareprodukt an. Mit Ablauf der Periode muß bei NichtVerlängerung des Lizenzvertrages die Software zurückgegeben werden.

Preisklasse: ca. DM 3000,- jährlich.

Vertrieb: SAS Institute GmbH, Neuenheimer Landstr. 28-30, 69120 Heidelberg.

LIMDEP

Der Name weist bereits auf einen wesentlichen Anwendungsbereich aus der Ökonometrie hin: »Limited Dependent Variable Models«. In diesem Sinn kann das Programm auch nicht als ein allgemeines Statistikpaket wie die o.g. Systeme bezeichnet werden. Es umfaßt aber eine sehr weite Klasse von Regressionsmodellen für sehr unterschiedliche Anwendungsbereiche, z.B. sehr ausgereifte Module zur Analyse von Zeitreihen, Ereignisdaten, Panel-Daten, diskrete Daten usf. mit zahlreichen Unteroptionen. Die Dokumentation ist sehr übersichtlich und gibt eine sehr differenzierte Erläuterung der über 50 implementierten Programme. Die bereits mitgelieferten Beispieldatensätze einschließlich der korrespondierenden Auswertungsdateien sind in dem Lehrbuch »Greene, W.H., 1990: Econometric analysis. London/New York: MacMillan« ausführlich abgehandelt.

Neben einer Vielzahl von Prozeduren aus dem Bereich der Regressionsanalyse/Ökonometrie bietet LIMDEP auch ausgezeichnete Möglichkeiten zur graphischen Darstellung von Modellevaluations-Statistiken, allerdings keine Business-Graphik im engeren Sinn (wie z.B. Kreis-/Balken- oder Liniendiagramme). Die Stärke liegt hier eher in den vielfältigen Plot-Varianten der Auswertungsprozeduren begründet. LIMDEP ermöglicht darüberhinaus die einfache Anwendung mathematischer Verfahren aus der Matrizenrechnung mit Hilfe eines differenzierten Matrix-Algebra-Zusatzprogramms, das über mehr als 30 Befehle verfügt. Eine weitere Besonderheit steht mit der Programmierung von sog. Makros zur Verfügung. Damit kann der Anwender in flexibler Weise bereits bestehende Statistik-Unterprogramme erweitern oder Verfahrensvarianten entwickeln.

Trotz der Verschiedenartigkeit und Komplexität der in LIMDEP integrierten Statistik-Programmanwendungen wahrt das Programm seinen Systemcharakter: Es benutzt eine einheitliche, innerhalb des Programms universell gültige Befehlsstruktur, die vom Anwender leicht und schnell erlernbar ist. Eine Benutzerführung mit Hilfe eines Menüsystems gibt es derzeit nicht, sieht man von den über die Funktionstasten steuerbaren Aufruf von Auswahloptionen ab. Die Programmsteuerung selbst erfolgt entweder interaktiv oder mit Hilfe eines Steuer-

Programms, das mit Hilfe eines implementierten Texteditors angelegt modifiziert und abgespeichert werden kann. Im interaktiven Modus erfolgt die Steuerung zeilenweise, bei der jede Programmzeile unmittelbar nach der Eingabe abgearbeitet wird (und somit im Falle einer fehlerhaften Eingabe auch direkt korrigierbar ist).

Bezüglich der Breite der verfügbaren regressionsanalytischen Verfahren zur Analyse unterschiedlicher Datenstrukturen ist LIMDEP das zur Zeit umfassendste System.

Preisklasse: ca. DM 1000,- (ca. 450 US \$).

Vertrieb: Econometric Software, Inc., 43 Maple Avenue, Bellport, NY 11713, USA.

Die tabellarische Übersicht (s. Tab. 1) faßt noch einmal die derzeit wichtigsten Statistik-Programmpakete zusammen, wobei wir Vorschläge für die möglichen Anwendergruppen mit einbeziehen. Unter »Spezialisten« verstehen wir nicht DV-Experten, sondern jene Anwender, die sehr spezielle multivariate Analyseverfahren einsetzen wollen (z.B. differenzierte Verfahren der Ereignisdatenanalyse, logistische Regressionsanalyse, loglineare Modelle, Zeitreihenanalyse etc.).

4. Vergleich der Statistik-Programmpakete

Unter der existierenden Software gibt es derzeit eine Reihe von leistungsfähigen Programmpaketen, die sich neben oder ergänzend zu dem in den Sozialwissenschaften weit verbreiteten Programmpaket SPSS-X nutzbringend verwenden lassen. Eine umfassende vergleichende Beurteilung würde sicherlich den Rahmen dieses Beitrages sprengen. Ein Gesamtbewertungsversuch einzelner, universeller Programmpakete hinsichtlich ihrer Leistungsmerkmale und eine nachfolgende Rangbildung ist ohnehin nicht möglich, da die einzelnen Produkte sich nur in ausgewählten Einzelaspekten unterscheiden. Die Abweichungen in der Leistungsfähigkeit lassen sich daher nur in einem Leistungsprofil abbilden. Laufende Programmverbesserungen im Rahmen neuer Updates erschweren darüberhinaus einen absolut gültigen Programmvergleich. Die Einordnung ausgewählter universeller Programmpakete in das Gesamtspektrum der möglichen Leistungsmerkmale beruht im folgenden auf einem Aggregat von Kennzeichen, das sich in folgende Teilklassen zerlegen läßt

- a) Allgemeine Systemcharakteristika:
- der Preis¹³,

¹³ Einige Programmpakete sind auch als Campus-Landeslizenzen zu einem erheblich günstigeren Preis erhältlich (z.B. SPSS/PC+ in NRW). Auskünfte erteilen die Beratungskräfte der örtlichen Hochschulrechenzentren.

Tab. 1: Überblick über die ausgewählten Statistik-Programmpakete und potentielle Anwendergruppen

Programmpaket	Preis	Anfänger	Fortgeschrittene	Spezialisten
SPSSx unter Windows	DM 2500,- (Basis)	ja	ja	ja
BMDP	ca. DM 3200,-	nein	ja	ja
CSS: STATISTICA unter Windows	ca. bis DM 2300,-	ja	ja	nein
ALMO	ca. DM 500,-	ja	ja	nein
NSDstat+	ca. DM 500,-	ja	nein	nein
P-STAT	ca. DM 3000,-	nein	ja	ja
SYSTAT	ca. DM 2000,-	nein	ja	ja
STATGRAPHICS PLUS	ca. DM 2600,-	ja	ja	nein
SAS	ca. DM 3000,- jährl. Leasinggebühr	nein	ja	ja
LIMDEP	ca. DM 1000,-	nein	nein	ja

Anmerkung: Der angegebene Preis ist in Einzelfällen von untergeordneter Bedeutung. Wenn eine Hochschullizenz für das jeweilige Programm am örtlichen Hochschulrechenzentrum erhältlich ist, ist der Preis meist erheblich niedriger.

- die Breite der unterstützten Hardware (Betriebssysteme),
- die Unterstützung eines arithmetischen Coprozessors,
- der Komfort der Benutzeroberfläche,
- die Programmierfähigkeit (Einbindung einer systemeigenen Programmiersprache),
- Online-Hilfefunktionen,
- Umfang und Verständlichkeit der Programmdokumentation,
- Lesbarkeit des Druckprotokolls,
- integrierte Gestaltbarkeit des Druckprotokolls.

b) Datenmanagement und Datenaustausch mit anderen Programmen:

- Größe der verarbeitbaren Datenmatrix,
- Breite der Datenaustauschformate,
- systemspezifischer Editor,
- Datentransformations- und -modifikations-Funktionen,
- Import-/Export-Schnittstellen.

c) Statistische Prozeduren:

Ein wichtiges Unterscheidungsmerkmal ist das jeweils gebotene Methodenspektrum und die Breite der verfügbaren, prozedurspezifischen Unteroptionen. Die unterschiedlichen statistischen Auswertungsverfahren lassen sich - je nach Zielsetzung - in folgende Großgruppen zusammenfassen:

- univariate Deskriptivstatistik (Mittelwert, Modus, Quantile, Median, kleinsten und größter Wert, Varianz, Standardfehler, Schiefe, Konzentration),
- Häufigkeitsverteilungen (univariate und bivariate Tabellierung, mehrdimensionale Kontingenztabellen, Assoziationsmaße),
- univariate und bivariate statistische Tests (parametrische Ein-/Zwei-Stichprobentests, nichtparametrische Tests, einfaktorielle Varianzanalyse),
- Korrelationsanalyse (Pearson's/Spearman's Korrelationskoeffizient, Korrelationsmatrix, statistischer Test für Korrelationen, partielle, multiple, kanonische Korrelation),
- Regressionsanalyse (einfache und schrittweise multiple Regression, Ridge-Regression, nichtlineare Regression, logistische Regression),
- Varianzanalyse (mehrfaktorielle Varianz- und Kovarianzanalyse, multivariate Varianz- und Kovarianzanalyse MANOVA/MANCOVA),
- Clusteranalyse (Aggregation einer Distanz-Matrix, hierarchische Verfahren, explorative Schnellverfahren),
- Faktorenanalyse (Hauptkomponentenmethode, Hauptfaktorenanalyse, alternative Schätzverfahren, Faktormatrix-Rotationsmethoden, Input-/Output-Optionen),
- Diskriminanzanalyse (lineare deskriptive Diskriminanzanalyse, quadratische Diskriminanzanalyse, schrittweise Diskriminanzanalyse, kanonische Analyse),
- Loglineare Modelle und Korrespondenzanalyse (hierarchische und nichthierarchische loglineare Modellbildung, Korrespondenzanalyse),

- Analyse diskreter abhängiger Merkmale (Sterbetafel-Analyse für verschiedene Verteilungen, Gruppenvergleich von Überlebensfunktionen, semiparametrische und parametrische Regressionsmodelle für Übergangsraten, Logit- und Probit-Modelle für Individualdaten),
 - Zeitreihenanalyse und Prognoserechnung (**deskriptive Methoden, Box-Jenkins-Analyse** ohne/mit automatischer Modellidentifikation, Fourier-Analyse, Autokorrelations-Analyse, etc.).
- d) Statistische Graphik:
- Business-Graphik (unterschiedliche Balken-, Linien-, Tortendiagramme),
 - Repräsentation einer Variablenstruktur (Box-Plots, **Lorenzkurve, Bar-Charts**),
 - Repräsentation einer Objektstruktur (**Chernoff-Gesichter, Hartigan-Bäume** etc.),
 - Zweidimensionale Scatter-Diagramme (Regressionsgerade, Punktwolke, Residuen, Konfidenzbereiche).

In dem Aufsatz »Statistical Software packages for PCs - a market survey« haben Bausch & Bankhofer 17 universelle Software-Produkte einem Produktvergleich zugrundegelegt und ihre Auswahl entlang o.g. Performance-Einzeldimensionen mit Hilfe einer sehr ausdifferenzierten Scoring-Methode in einem Praxis-Test beurteilt. Die Tabellen 2a und 2b enthalten einen Ausschnitt aus den wichtigsten Analyseergebnissen.

Ein Scoring-Tabellenwert von 1.00 bedeutet daß das korrespondierende Programmpaket sämtliche Anforderungen in den genannten Einzeldimensionen perfekt erfüllt; ein Wert von 0 repräsentiert dagegen eine sehr schwache Implementierung der genannten Leistungsmerkmale oder indiziert ein völliges Fehlen. Die Autoren haben die Einzelergebnisse abschließend mit Hilfe clusteranalytischer Verfahren analysiert, um zu einem Gesamtergebnis zu kommen. Dabei ergaben sich u.a. folgende Großgruppen von Programmpaketen:

Gruppe 1: BMDP, SAS, SPSS bzw. SPSS/PC+, SYSTAT; Allround-Pakete, die unter verschiedenen Betriebssystemen angeboten werden; zahlreiche Datenaustauschformate und Transformationsfunktionen, breite Auswahl von Statistikprozeduren. Die Graphik ist dagegen - falls überhaupt implementiert - eher dürftig (Ausnahme: SYSTAT, SAS).

Gruppe 2: CSS, STATGRAPHICS, SPSS unter Windows; Graphik-Experten, die nur für PCs erhältlich sind; zahlreiche Datenaustauschformate, breite Auswahl von Statistikprozeduren und Graphik-Tools.

Gruppe 3: P-STAT; Individualist; zahlreiche Datenaustauschformate, breite Auswahl von Statistikprozeduren, lediglich elementare Graphik.

Eine Analyse der Tabelle mittels einer Hauptkomponentenanalyse mit Varimax-Rotation führt zu zwei Faktoren, die als »Statistik-Komponente« und

Tab. 2a: Überblick über die Leistungsmerkmale der wichtigsten Statistik-Programmpakete.

Performance-Merkmale	B M D P	C S S	P S T A T	S A S
Preis	1.00	0.50	1.00	1.00
unterstützte Hardware	0.20	0.00	0.60	0.60
Coprozessor Unterstützung	1.00	1.00	1.00	1.00
Komfort der Benutzeroberfläche	1.00	1.00	1.00	1.00
Programmierungsfähigkeit	1.00	0.75	1.00	1.00
Online-Hilfe	0.63	0.55	0.84	0.90
Programmdokumentation	0.76	0.55	0.54	0.97
Lesbarkeit des Druckprotokolls	0.75	0.50	0.25	0.75
Gestaltbarkeit des Druckprotokolls	0.90	0.90	0.00	1.00
Datenmatrixgröße	1.00	0.83	0.86	1.00
Datenaustauschformate	1.00	0.85	0.85	1.00
Editor	0.10	0.93	0.13	0.70
Datentransformation	0.75	0.41	0.81	0.74
Import/Export-Schnittstellen	0.78	0.88	0.95	0.68

Quelle: Bausch, T. / Bankhofer, U., 1992: Statistical software packages for PCs - a market survey, in: Statistische Hefte 33, 283-306.

(Anmerkung: Die Spalte "SPSS unter Windows" haben wir nachträglich gebildet.)

Tab. 2a: Fortsetzung

Performance-Merkmale	B M D P	C S S	P S T A T	S A S
univariate Deskriptivstatistik	0.97	0.80	0.87	0.87
Häufigkeitsverteilung	0.96	0.86	0.96	0.97
uni-und bivariate statistische Tests	0.56	0.64	1.00	0.78
Korrelationsanalyse	0.85	0.75	0.95	0.85
Regressionsanalyse	0.86	0.71	0.34	0.78
Varianzanalyse	1.00	0.65	0.32	1.00
Clusteranalyse	0.46	0.58	0.03	0.63
Faktorenanalyse	0.83	0.74	0.58	0.87
Multidimensional scaling	0.00	0.40	0.00	0.00
Diskriminanzanalyse	0.25	0.65	0.20	0.95
Loglineare Modelle	0.75	0.25	0.25	0.75
Logistische Regression/Ereignis- datenanalyse	0.66	0.83	0.39	1.00
Zeitreihenanalyse	0.80	0.78	0.58	0.92
Business-Graphik	0.00	1.00	0.46	1.00
Repräsentation einer Variablen- struktur	0.40	0.80	0.65	0.68
Repräsentation einer Objektstruktur	0.10	0.90	0.00	0.00
Zweidimensionale Scatter-Dia- gramme	0.30	0.62	0.24	0.57

Tab. 2b: Überblick über die Leistungsmerkmale der wichtigsten Statistik-Programmpakete.

Performance-Merkmale	S P S S / W I N	S P S S	S T A T I S T I C S	S Y S T E M
Preis	1.00	1.00	0.75	0.75
unterstützte Hardware	0.60	0.60	0.00	0.60
Coprozessor Unterstützung	1.00	1.00	1.00	1.00
Komfort der Benutzeroberfläche	1.00	1.00	1.00	1.00
Programmierungsfähigkeit	0.75	0.75	0.00	0.75
Online-Hilfe	0.76	0.76	0.52	0.47
Programmdokumentation	0.78	0.78	0.85	0.81
Lesbarkeit des Druckprotokolls	0.25	0.25	0.75	0.50
Gestaltbarkeit des Druckprotokolls	1.00	1.00	1.00	1.00
Datenmatrixgröße	0.97	0.97	0.62	0.84
Datenaustauschformate	1.00	1.00	1.00	1.00
Editor	1.00	1.00	0.93	0.60
Datentransformation	0.50	0.50	0.22	0.75
Import/Export-Schnittstellen	0.80	0.80	0.83	0.83

Quelle: Bausch, T. / Bankhofer, U., 1992: Statistical software packages for PCs - a market survey, in: Statistische Hefte 33, 283-306.

(Anmerkung: Die Spalte "SPSS unter Windows" haben wir nachträglich gebildet.)

Tab. 2b: Fortsetzung

Performance-Merkmale	S P S / W I N	S P S S	S T A T I S T I C S G R A P H	S Y S T E M A T
univariate Deskriptivstatistik	0.88	0.88	0.83	0.77
Häufigkeitsverteilung	0.95	0.95	0.92	0.88
uni-und bivariate statistische Tests	0.86	0.86	0.78	0.64
Korrelationsanalyse	0.85	0.85	0.85	0.95
Regressionsanalyse	0.54	0.54	0.43	0.80
Varianzanalyse	0.66	0.66	0.40	0.99
Clusteranalyse	0.54	0.54	0.37	0.47
Faktorenanalyse	0.87	0.87	0.71	0.71
Multidimensional scaling	0.80	0.00	0.00	0.40
Diskriminanzanalyse	0.65	0.65	0.20	0.65
Loglineare Modelle	1.00	1.00	0.25	0.25
Logistische Regression/Ereignisdatenanalyse	1.00	0.74	0.09	0.63
Zeitreihenanalyse	0.82	0.82	0.70	0.72
Business-Graphik	1.00	0.60	0.97	0.89
Repräsentation einer Variablenstruktur	0.40	0.40	0.80	0.77
Repräsentation einer Objektstruktur	0.90	0.10	0.30	0.90
Zweidimensionale Scatterdiagramme	0.33	0.33	0.58	0.94

»Graphik-Komponente« interpretiert werden können. Danach ergibt sich folgende Rangfolge in den Programmpaketen:

Statistik-Komponente:

- (1) SAS
- (2) BMDP
- (3) SPSSx, SPSS/PC+
- (4) P-STAT

Graphik-Komponente:

- (1) SYSTAT
- (2) STATGRAPHICS
- (3) CSS

Eine Ausnahme bildet die von uns zusätzlich berücksichtigte SPSS-Version unter Windows; diese jüngste Entwicklung von SPSS Inc. ist sowohl in der Breite der **Statistik**-Prozeduren als auch in der Neuaufnahme von integrierten, vielfältigen Business-Graphikoptionen verbessert worden. Als (vorläufiges) Fazit fassen Bausch & Bankhofer ihren Vergleichstest zusammen: »The existence of a global best Statistical Software package cannot be indicated by the results of the tests ... Without a profile of the potential users System requirements it is not possible to find the product fitting these requirements best« (a.a. O., S. 304). Die Tests können nur eingeschränkt zur Beurteilung der Gesamtleistungsfähigkeit der Systeme herangezogen werden, da sie spezielle Leistungsaspekte in unterschiedlichen Preissegmenten betrachten. In dem unteren Preissegment dürfte das Programm CSS einen guten Platz einnehmen. SYSTAT ist führend in dem mittleren Preissegment. In dem oberen Preissegment sind sowohl SAS als auch SPSS unter Windows führend. Lediglich SAS hat im Hinblick auf den Benutzerkomfort leichte Defizite.

Der Vergleich von universellen Programmpaketen mit ihren vielfältigen Statistikprozeduren darf letztlich nicht darüber hinwegtäuschen, daß für spezielle Analyseprobleme einzig die sog. »Spezialisten« befriedigende Prozeduren bzw. hinreichend differenzierte Einzeloptionen anbieten.

Ausblick

Seitdem in den letzten Jahren nunmehr auch die Software (vor allem die Betriebssysteme) sich der inzwischen schon immensen Leistungsfähigkeit der PC-Hardware angeglichen haben, ist davon auszugehen, daß in fast allen Fällen (auch bei großen, hierarchisch aufgebauten Datensätzen) handelsübliche PCs ohne weiteres 'Tuning' und spezielle - nur Insidern bekannten Tricks - die ihnen für die statistische Analyse von historischen (Massen-)Daten gestellten Aufgaben bewältigen können. Mit der Verfügbarkeit von CD-ROMs als Massenspeicher, sowie mit der verstärkten Anbindung der einzelnen PCs an Computernetzwerke dürften auch Engpässe bei der Speicherung kein unüberwindbares Hindernis mehr darstellen.

Es ist abzusehen, daß die Benutzung von PCs sich bald vollständig als Standard durchgesetzt haben wird. Daher wird vieles in der zu erwartenden Ent-

Wicklung an den erweiterten Möglichkeiten zu messen sein, die hier die Technologie im Bereich der Anwendungssoftware bietet. Mit den 32-bit Betriebssystemen, den Multitasking-Fähigkeiten (z.B. unter den Betriebssystemen OS/2 oder UNIX) wird sich die Leistungspalette erweitern. Bei den am Industriestandard angelehnten Softwareprodukten wird sich durch neue optische Möglichkeiten in der Bildschirmgestaltung und durch verbesserte Bedienungsvorteile (graphische Benutzeroberflächen) auch die Anwendung wesentlich vereinfachen. Mithin wird auch der einzelne Anwender immer weniger Zeit mit dem Erlernen von Programmpaketsprachen (d.h. der programmspezifischen Syntax) verbringen müssen. Selbst die Arbeiten mit Betriebssystemsteuerungen verringern sich infolge der inzwischen immer komfortabler werdenden Benutzeroberflächen. Bei den Programmpaketen insgesamt scheint sich eine weitestgehende Konvergenz anzudeuten, die auf eine, oberflächlich betrachtet, einfachere Architektur und Bedienbarkeit hinauslaufen.

Wenngleich die führenden Hersteller wie SAS und SPSS zeitweilig ihr Interesse deutlich stärker auf den freien Markt als auf die universitäre Forschungslandschaft ausrichteten, so stellen sie - dessen ungeachtet - heute immer noch wichtige Instrumente der statistischen Datenanalyse dar. Den Paketen SYSTAT (Stärke: Visualisierung) und LIMDEP (spezielle Analyseprozeduren für diskrete abhängige Merkmale, einfache Handhabung) könnte u.U. in Zukunft eine größere Rolle zukommen. Das Programm BMDP hat in den vergangenen Jahren zunehmend an Bedeutung verloren, zumindest in dem Bereich der sozialwissenschaftlichen Anwendungen. Hier zeichnet sich ein Rückzug auf die ursprünglichen Anwendungsfelder der Biostatistik (bzw. Biomedizin) ab.

5. Wissensbasierte Systeme in der Statistik

In den zurückliegenden Jahren nahm auch in den Sozialwissenschaften der Kreis der Anwender mathematisch-statistischer Methoden eine explosionsartige Entwicklung. Bedingt wurde dieser Prozeß durch die rasante Entwicklung der Rechentechnik und entsprechend verfügbarer statistischer Software. Die daraus resultierenden neuen Probleme führten unter anderem folgerichtig zur Herausbildung des neuen Fachgebiets *Informatik in der Statistik* (engl.: computational statistics«). Als interdisziplinäres Forschungsgebiet umfaßt die statistische Informatik die weitgehend unabhängigen Fachrichtungen der mathematischen Statistik, der Biometrie und der Informatik. In das Fachgebiet von *computational statistics* gehört auch die Entwicklung und Bewertung von *statistischen Expertensystemen*.

Der anwendungsorientierte Alltag bei der Bearbeitung statistischer Probleme ist dadurch bestimmt, daß überwiegend die großen statistischen Auswertungssysteme wie SPSS, BMDP, SAS etc. zum Einsatz gelangen. Aber auch bei Einsatz dieser Auswertungssysteme werden häufig noch statistische Verfahren

angewandt, die die Fragestellung unter Umständen nur unzureichend widerspiegeln, obwohl vielfach innerhalb des verwendeten Auswertungssystems besser geeignete Verfahren zur Verfügung stehen. Zum Beispiel werden häufig parametrische Verfahren auch auf solche Fälle angewandt, in denen die Annahmen der diesen Verfahren zugrunde liegenden mathematischen Modelle der Fragestellung weniger gut entsprechen als die Annahmen nichtparametrischer Verfahren.

Ein vielversprechender Versuch, die Fehleranfälligkeit bei der Arbeit mit den großen Systemen zu reduzieren, besteht zum einen in der ständigen Verbesserung der Mensch-Computer-Schnittstelle und zum anderen in dem Einsatz von *Methoden der künstlichen Intelligenz* (KI)¹⁴, die zur Entwicklung von *Wissensbasierten Systemen* führte. Unter Wissen soll hier vorerst ganz allgemein die Verfügbarkeit von Informationen über Fakten, Konzepte und Methoden aus einem bestimmten Problembereich und eine geeignete symbolische Repräsentation mittels Objekten, Fakten und Regeln in operationaler Form verstanden werden.

Eine Klassifikation des menschlichen Wissens kann in informelles, technisches und formales Wissen vorgenommen werden. Das informelle Wissen ist dadurch gekennzeichnet, daß der Mensch es durch Zuschauen und Nachahmen erwirbt. Es handelt sich hierbei nicht um Expertenwissen. Da dieses Wissen nicht formalisiert ist, macht seine softwaretechnische Implementierung enorme Schwierigkeiten. Für die maschinelle Verarbeitung ist die Unterscheidung zwischen technischem und formalem Wissen von gravierender Bedeutung. Zwar ist beides formalisiert, d.h. als Gerüst von *Fakten* und *Regeln* darstellbar, wesentlich ist aber, woher diese Formalisierung stammt und damit auch, wie und in welcher Form das Wissen repräsentiert ist und welche Wissensänderungen zu erwarten sind. *Technisches Wissen* ist in der Regel aus Theorien abgeleitet. Als Beispiel seien die gesamte Mathematik, physikalische Formeln und rechen-technische Verfahren genannt. Auch die üblichen Programmiersprachen sind dem technischen Wissen zuzuordnen. Sie beschreiben Algorithmen, d.h. Rechenvorschriften, und erlauben es damit, in der Realität vorgefundene Systeme auf einem Rechner zu modellieren. *Formales Wissen* läßt sich nicht in numerischen Formeln, sondern z.B. verbal und formal in wenn-dann-Aussagen darstellen.

Wissensbasierte Software-Entwicklung, die flogranntiertechnik der Expertensysteme, stellt für die Verarbeitung formalen Wissens ebenso bessere Hilfsmittel bereit, wie die strukturierte Programmierung dies für die softwaretechnische Bearbeitung technischen Wissens getan hat. Unter einem Expertensystem soll hier die Einführung einer auf Wissen basierenden Komponente in einer bestimmten Software verstanden werden. Diese Software bietet einem

¹⁴ Die künstliche Intelligenz ist ein Teilbereich der Informatik, in dem versucht wird, die kognitiven Leistungen des Menschen zu verstehen und mit Hilfe der EDV bzw. darauf laufenden Programmen nachzuvollziehen.

Nutzer eine entsprechende intelligente Beratung oder trifft eine intelligente Entscheidung bezüglich einer Verarbeitungsfunktion.

Die vielleicht wichtigste Eigenschaft eines Expertensystems ist, daß es neben der Datenbasis eine *Wissensbasis* aufweist, in welcher möglichst alle Aspekte des benötigten Fachwissens in änderungsfreundlicher Form gespeichert werden. Das Wissen umfaßt Fakten - im Sinne von allgemein als geltend akzeptierten Fachinformationen - und Heuristiken, die man als »Faustregeln« bezeichnen könnte, die sich aus Erfahrungen in einem Anwendungsbereich ergeben haben. Durch einfache Erweiterung der Wissensbasis können die Systeme an neue Gegebenheiten angepaßt werden. Der Wissens- und damit der Anwendungsbereich eines solchen Systems ist immer sehr eng auf ein bestimmtes Problemfeld begrenzt, in dem das System Hilfen zur Problemlösung geben soll. Es hat sich herausgestellt, daß die Systeme umso erfolgreicher sind, je enger dieses Problemfeld begrenzt wird.

Der »Problemlösungsprozeß« mußte - unabhängig von der Aufgabe - in der Vergangenheit vor allem von Menschen getragen werden. Der Computer konnte nur als Hilfsmittel zur Informationsbereitstellung verwendet werden. Die Abbildung 3 verdeutlicht, wie dies durch Wissensbasierte Systeme in Zukunft verändert werden soll.

Die Hilfen, die die klassische EDV bisher dem Menschen im Entscheidungsprozeß lieferte, muß nun den Wissensbasierten Systemen zur Verfügung gestellt werden. Das heißt, wissensbasierte Systeme müssen auf Daten und Routinen der konventionellen DV zugreifen können. Die beiden Welten können nicht isoliert nebeneinander bestehen, sondern sie müssen zusammenwachsen. Die Abbildung 4 skizziert in allgemeiner Form die Technologie Wissensbasierter Systeme.

Die mit dem Aufkommen der Datenverarbeitung zur Verfügung stehenden Anwendungsprogramme führten noch alle benötigten Verarbeitungsdaten mit sich, die weitere Entwicklung führte zur Trennung von Befehls- und Verarbeitungsdaten, so daß mehrere Programme auf denselben Datenbestand zugreifen konnten. Im nächsten Schritt - der Entwicklung der Wissensbasierten Systeme - ging man konsequenterweise einen Schritt weiter der Trennung von Anwendungsdaten, fachspezifischen Regeldaten und Befehlsdaten (s. d. Abbildung 5).

Das Fachwissen wird in der Wissensbasis verankert. Sinn dieser Vorgehensweise ist es, veränderliches oder ständig sich erweiterndes Wissen in der Wissensbasis ändern und korrigieren zu können, ohne damit den Programmablauf zu beeinflussen, also ohne gleichzeitige Programmänderung. Wissensbasierte Systeme sollen herkömmliche Anwendungen nicht ersetzen, sondern ergänzen.

Unter Statistik als Anwendungsgebiet von Wissensbasierten Systemen wird im folgenden nicht die mathematische Statistik verstanden, sondern vor allem die angewandte Statistik. Wissensbasierte Systeme in der Statistik lassen sich dann als Weiterentwicklungen der statistischen Auswertungssysteme ansehen. Neben einer für den Anwender möglichst einfach aufzurufenden statistischen

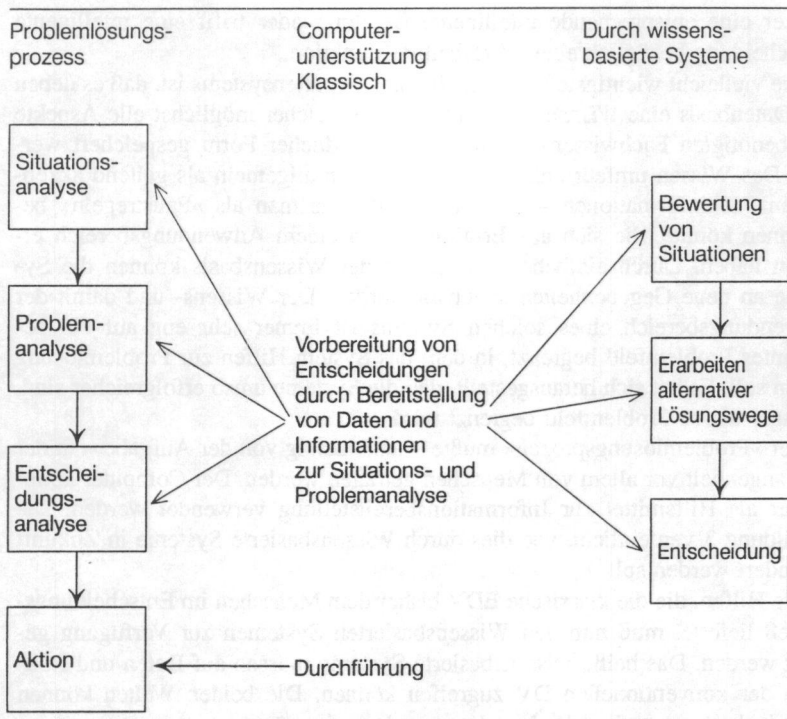


Abb. 3: Wissensbasierte Systeme im Problemlösungsprozeß
 (Quelle: Maciejewski, P.G., 1989: Künstliche Intelligenz. Heidelberg, S. 17).

Datenanalyse sollen solche Systeme künftig in verstärktem Maße sicherstellen, daß die verwendete statistische Methode und das gewählte statistische Modell für das zu untersuchende Problem auch angemessen ist. Die Abbildung 6 illustriert unter einer idealisierten Betrachtungsweise den Aufbau Wissensbasierter Systeme. Es handelt sich dabei um eine Erweiterung der Architektur von Daten- und Programmbanksystemen. Hinzugekommen ist die Wissensbank, gegliedert in Wissen über Daten und über Programme bzw. Methoden. In der Datenbank sind z.B. Angaben über die Skalierung und die Verteilung von Merkmalen enthalten, in der Wissensbank Wissen über statistische Modelle und Verteilungsannahmen von statistischen Testverfahren. Das Daten- und Programmbankverwaltungssystem ist ergänzt um ein Wissensverwaltungssystem; die Schnittstellen sind entsprechend erweitert. Das Wissensbankverwaltungssystem verwendet - nach bestimmten Vorgaben - das Wissen aus der Wissensbank zur Problemlösung aufgrund von Daten aus der Datenbank. Wichtige

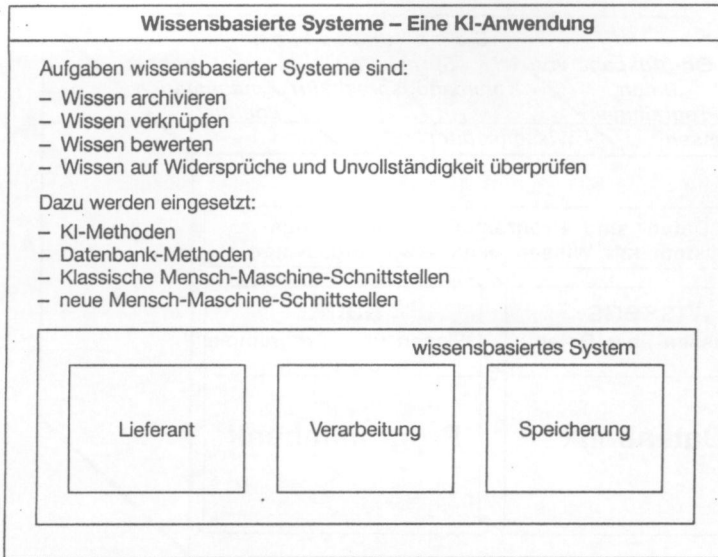


Abb. 4: Wissensbasierte Systeme

(Quelle: Maciejewski, P.G., 1989: Künstliche Intelligenz. Heidelberg, S. 3).

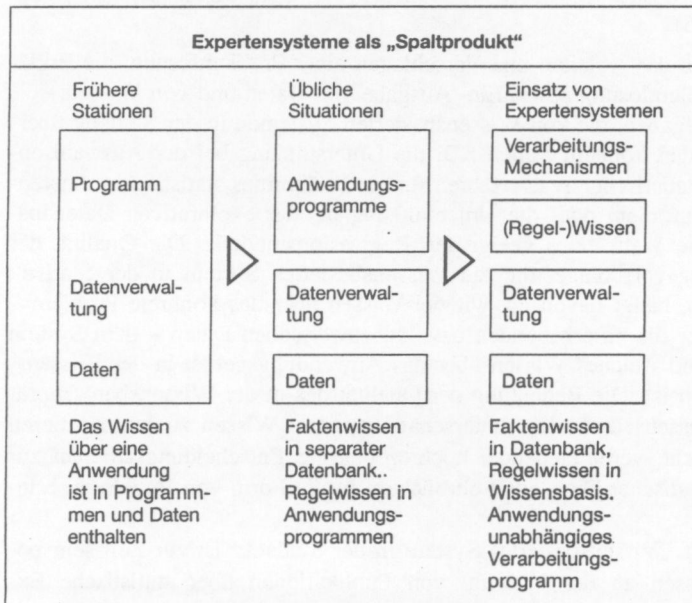


Abb. 5: Evolution von Softwarearchitekturen

(Quelle: Maciejewski, P.G., 1989: Künstliche Intelligenz. Heidelberg, S. 32).

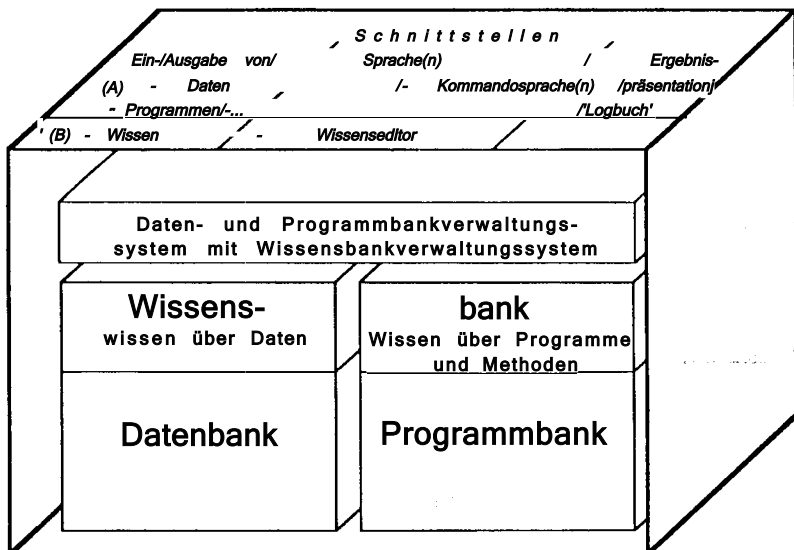


Abb. 6: Aufbau eines wissensbasierten Systems (A: Anwenderschnittstellen; B: Wissensverwalterschnittstellen).

(Quelle: Haag, U./Haux, R./Kieser, M., 1992: Statistische Auswertungssysteme. Stuttgart/Jena, S. 35).

Aufgaben sind das Ableiten und Vorschlagen einer Problemlösung, die Erklärung der Problemlösung sowie Ein-/Ausgabe von Daten und von Wissen.

Zu den Aufgaben, die von Wissensbasierten Systemen in der Statistik übernommen werden können, zählen z.B. die Unterstützung bei der Auswahl angemessener statistischer Testverfahren für ein bestimmtes statistisches Anwendungsgebiet/-problem oder die Unterstützung bei der explorativen Datenanalyse durch die Wahl eines geeigneten Regressionsmodells. Die Qualität der Problemlösungsvorschläge, die ein Wissensbasiertes System in der Statistik abgeben kann, hängt davon ab, wieviel Wissen über das konkrete Problem - also etwa über die zu erhebenden bzw. auszuwertenden Daten - dem System bekannt ist und welches Wissen über das Anwendungsgebiet in der Wissensbank enthalten ist. Die Bedeutung der Qualität des in der Wissensbank repräsentierten Wissens darf nicht unterschätzt werden. Wissen zu repräsentieren heißt z.B. nicht vereinfachende, kochrezeptartige Entscheidungstabellen zur Auswahl statistischer Tests aus Lehrbüchern in die Form von Regeln zu bringen.

Der Begriff 'Wissensbasiertes System in der Statistik' ist zur Zeit sehr populär. Gemessen an der Vielzahl von Publikationen über statistische Expertensysteme ist die Zahl der real existierenden Systeme erstaunlich gering. Obwohl keines der Systeme im kommerziellen Sinne als ausgereift gelten kann,

lassen sich trotzdem bereits verschiedene Konzepte hinsichtlich Aufbau der Wissensbasis und Gestaltung der Benutzer-Schnittstelle erkennen. Die hier diskutierten Systeme sind eher prototypische Versionen oder solche, die zwar lauffähig sind, die jedoch nur in ersten, einfachen Grundaufbaustufen existieren

Vier solcher Systeme, CADEMO, GLIMPSE, PANOS und REX sollen im folgenden kurz beschrieben werden.

CADEMO

CADEMO steht für Computer Aided Design of Experiments and Modeling. Das System wurde unter der Federführung von D. Rasch an der Akademie der Landwirtschaftswissenschaften der ehemaligen DDR in Dummerstorf (bei Rostock) entwickelt und ist für PCs erhältlich. Zielsetzung ist die Entscheidungsunterstützung bei der experimentellen Studienplanung, insbesondere für konfirmatorische Analysen in der Biometrie. Derzeit liegt der Schwerpunkt bei der Unterstützung statistischer Methoden, die in der Landwirtschaft üblich sind; das System wird allerdings weiterentwickelt. Derzeit ist für spezifisch sozialwissenschaftliche Problemstellungen das implementierte Methodenspektrum zu eingengt.

GLIMPSE

GLIMPSE steht für »**GLIM** + Prolog + Statistical Expertise«. Es wurde unter maßgeblicher Beteiligung des Statistikers J.A. Neider am Imperial College of Science und Technology, London, als komfortable Anwenderschnittstelle für das statistische Auswertungssystem GLIM (»Generalized Linear Interactive Modelling«) entwickelt. Es ist für Rechner des Typs SUN 3 erhältlich (NAG, UNIX Workstation).

Ziel des Systems ist die komfortable und intelligente Unterstützung des Benutzers bei der Auswahl eines zu seinen Daten passenden linearen Modells. Der Anwender, der zumindest Grundwissen über verallgemeinerte lineare Modelle besitzen sollte, wird bei der Auswahl eines den Daten adäquaten verallgemeinerten linearen Modells unterstützt. Durch die Wahl eines entsprechenden Qualifizierungsgrades kann der Benutzer den Grad der Unterstützung durch das System bestimmen. Nach Überprüfung der spezifizierten Kommandos auf ihre semantische und syntaktische Korrektheit durch das System wird das Programm GLIM aktiviert, welches die eigentliche numerische Datenanalyse durchführt. Trotz zahlreicher Vereinfachungen durch die interaktive Benutzerführung wendet sich dieses System eher an statistische Experten. Das Methodenspektrum ist sehr weit gefächert, so daß auch sozialwissenschaftliche Problemstellungen weitgehend mit diesem System gelöst werden können.

PANOS

PANOS steht für »Parametric And Nonparametric Statistics«. Entwickelt wurde das System im Rahmen einer Dissertation von K.M. Wittowski an der Universität Tübingen¹⁵. PANOS existiert zur Zeit als Prototypversion auf einem Arbeitsplatzrechner und ist in BASIC implementiert.

PANOS ist im deutschsprachigen Raum das Pioniersystem unter den Expertensystemen zur Statistik. Der Anwender - vorausgesetzt er hat fundierte statistische Kenntnisse - soll unterstützt werden bei der Auswahl geeigneter statistischer Modelle. Darüberhinaus dient es als Assistenzsystem zur Formulierung der Testprobleme in dem Modell bzw. bei der Auswahl der geeigneten Teststatistik. Das System umfaßt eine breite Klasse parametrischer und nicht-parametrischer Tests. Die Daten können in PANOS gehalten werden. Das Datenbankschema enthält Merkmale sowie Angaben über diese Merkmale (z.B. das Skalenniveau). Für eingeschränkte sozialwissenschaftliche Problemstellungen ist das System geeignet.

REX

REX steht für »Regression Expert« und wurde von W.A. Gale und D. Pregibon in den Bell Laboratorien von AT&T entwickelt¹⁶. Das System ist zwar erhältlich, allerdings ist eine sehr dezidierte Hard- und Softwareumgebung erforderlich. Die Autoren betonen den rein experimentellen Charakter der Systementwicklung. Es handle sich auf keinen Fall um ein kommerziell vertreibbares System.

Zielsetzung ist die Unterstützung statistischer Laien, aber auch statistischer Experten, bei der Anpassung linearer Regressionsmodelle an gegebene Daten. Die Benutzerschnittstelle von REX wird durch natürlichsprachlichen Dialog und fortgeschrittener Window-Technik bestimmt. Die Entwicklung von REX ist abgeschlossen. In den Bell Laboratorien wurde in der Nachfolge an dem System STUDENT gearbeitet.

Zusammenfassung

Hinsichtlich der Konstruktionsprinzipien bei der Entwicklung von Expertensystemen in der Statistik lassen sich zur Zeit zwei prinzipielle Trends registrieren. Erstens werden um größere, abgeschlossene Auswertungssysteme mittels immer intelligenterer Nutzerschnittstellen »Expertenkomponenten« konstruiert (z.B. GLIMSE), die den Einsatz dieser Systeme komfortabler und hinsichtlich der Methodenauswahl stabiler machen. Zweitens werden kleinere Systeme

¹⁵ Wittowski, K.M., 1985: Ein Expertensystem zur Datenhaltung und Methodenauswahl für statistische Auswertungen. Dissertation, Universität Stuttgart.

¹⁶ Vgl. Gale, W.A. (ed.), 1986: Artificial intelligence and statistics. Reading: Addison-Wesley.

(z.B. PANOS) zu einem gut abgrenzbaren Methodenspektrum implementiert, welches eine weitgehende Strukturierung des Wissens über dieses statistische Gebiet gestattet. Ziel aller vorgestellten Systeme ist, die Anwender eines Wissensbasierten Systems durch die Reduzierung der Gefahr falscher Auswertungen zu unterstützen bzw. sicherzustellen, daß statistische Modelle problemadäquat angewandt werden.

Zusammenfassend sei festgestellt, daß der Anspruch an Wissensbasierte Systeme in der Statistik, zum einen Experten zu ersetzen und zum anderen Expertenwissen über ein so großes Teilgebiet wie der Statistik zu umfassen, von den derzeit verfügbaren Systemen nicht eingelöst wird. Es stellt sich auch grundsätzlich die Frage, ob diese Anforderungen überhaupt von wissensbasierten Systemen erfüllt werden können. Bei der praktischen Arbeit im Zusammenhang mit wissensbasierten Systemen in der Statistik wird auch künftig der unter einem allumfassenden Anspruch initialisierte Ansatz von Expertenwissen keine Rolle spielen.

Bibliographie

1. Übersichten:

- Birch, D7Denley, P., 1993: A historical computing bibliography. Version 1.0. Association for HSC and Humanities Computing Centre, Queen Mary & Westfield College. London. (Als Datenbank via FTP frei zugänglich).
- Brent, E.E. Jr./Anderson, R.E., 1990: Computer applications in the social sciences. Philadelphia.
- Deegan, M./Lee, S. (eds.), 1991: CTI centre for textual studies studies resources guide. Oxford.
- Faulbaum, F./Uehlinger, M. (Hrsg.), 1988: SOFTSTAT '87. Fortschritte der Statistik-Software 1. 4. Konferenz über die wiss. Anwendung von Statistik-Software, Heidelberg 1987. Stuttgart u.a.: G. Fischer.
- Faulbaum, F. u.a. (Hrsg.), 1990: SOFTSTAT '89. Fortschritte der Statistik-Software 2. 5. Konferenz über die wiss. Anwendung von Statistik-Software, Heidelberg 1989. Stuttgart u.a.: G. Fischer.
- Faulbaum, F. u.a. (ed.), 1992: SOFTSTAT '91. Advances in Statistical software 3. The 6th conference on the scientific use of statistical software, Heidelberg 1991. Stuttgart u.a.: G. Fischer.
- Faulbaum, F. (ed.), 1994: SOFTSTAT '93. Advances in statistical software 4. The 7th conference on the scientific use of statistical software, Heidelberg 1993. Stuttgart u.a.: G. Fischer.
- Fitch, N.E., 1980: Statistical and mathematical methods for historians: An annotated bibliography of selected books and articles, in: Historical methods, 13, S. 221-231.

- Francis, I., 1981: Statistical software - A comparative review. Amsterdam: North-Holland.
- Grossbart, S.R., 1992: Quantification and social science methods for historians: An annotated bibliography of selected books and articles, in: Historical Methods, 25, S. 100-120.
- Haag, U./Haux, R.V./Kieser, M., 1992: Statistische Auswertungssysteme. Eine Einführung in ihre Anwendung. Stuttgart u.a.: G. Fischer.
- Hall, L.D./Marshall, K.P., 1992: Computing for social research. Practical approaches. Belmont, Calif.: Wadsworth.
- Hausmann, F. u. a. (Hrsg.), 1987: Datennetze für die Historischen Wissenschaften, Probleme und Möglichkeiten bei Standardisierung und Transfer maschinenlesbarer Daten. Graz.
- Heise, B., 1994: Computer und Statistik. Bonn: Addison-Wesley.
- History and Computing (Software Reviews).
- History Micro Review (Hardware/Software Reviews).
- Lancashire, I. (ed.), 1991: The humanities computing yearbook 1989/90: A comprehensive guide to software and other resources. Oxford.
- Matsuba, R./Wardley, P., 1990: Computer application in the humanities: A reading list, in: Canadian Humanities Computing, 4, S. 1-8.
- Merkel, B./Sperling, B., 1990: Statistik für Mikrocomputer. Stuttgart u.a.: G. Fischer.
- Middleton, R./Wardley, P., 1990: Annual review of information technology developments for economic and social historians, in: Economic History Review.
- Spaeth, D.A., 1991: A guide to software for historians. Glasgow: Univ. Press.
- Welford, R. (ed.), 1990: Information technology for social sciences. Shipley.
- Wittenberg, R., 1993: Grundlagen computerunterstützter Datenanalyse. Handbuch für computerunterstützte Datenanalyse Bd. 1. Stuttgart u.a.: G. Fischer (UTB)
- Woodward, W. u.a., 1988: Directory of statistical microcomputer software. New York: Dekker.

2. Datenanalyse mit speziellen Statistik-Programmpaketen

2.1 SPSS (»Statistical Product and Service Solutions«, früher: »Statistical Package for the Social Sciences«)

- Babbie, E.V./Halley, F., 1993: Adventures in social research: Data analysis using SPSS. Newbury Park: Sage.
- Bauer, F., 1986: Datenanalyse mit SPSS, 2., rev. A. Berlin: Springer.
- Brosius, G., 1991: SPSS/PC+ Basics and graphics, 2. A., deutsche Ausgabe. New York: McGraw-Hill.
- Brosius, G., 1989: SPSS/PC+ Advanced statistics and tables, deutsche Ausgabe. New York: McGraw-Hill.

- Frenzel, G./Hermann, D. (Hrsg.), 1989: Statistik mit SPSSx: Eine Einführung nach M.J. Norusis. Stuttgart u.a.: G. Fischer.
- Friede, CySchirra-Weinrich, L., 1991: Standardsoftware Statistische Datenanalyse SPSS/PC+. Reinbek: Rowohlt.
- Hedderston, L., 1991: SPSS-PC+ made simple, 2nd ed. Belmont, Calif.: Wadsworth.
- Hermann, D. (Hrsg.), 1994: SPSS/PC+ Benutzerhandbuch. Band 2: Komplexe statistische Verfahren, Tabellenanalyse und graphische Analyseverfahren. Stuttgart: G. Fischer.
- Kaehler, W.-M., 1990: Statistische Datenanalyse mit SPSS-PC+: Eine Einführung in Grundlagen und Anwendung. Wiesbaden: Vieweg.
- Kinnear, P.R./Gray, C.D., 1992: SPSS-PC+ made simple. Hove u.a.: Erlbaum.
- Klemm, E., 1992: Dateneingabe und Datenmanagement in SPSS-PC+. Mannheim: BI.
- Klemm, E., 1994: Computerunterstützte Datenerfassung. Stuttgart: G. Fischer.
- Levine, G., 1991: A guide to SPSSx for analysis of variance. Hillsdale, NY: Erlbaum.
- Matiasko, W., 1990: Statistische Datenanalyse mit Mikrocomputern. Einführung in P-STAT und SPSS/PC. München/Wien: Oldenbourg.
- Pfeifer, A., 1991: Datenanalyse mit SPSS-PC+. München/Wien: Oldenbourg.
- Rowland, D./Arkkelin, D./Crisler, L., 1991: Computer based data analysis: Using SPSSx in the social and behavioral sciences. Chicago: Nelson-Hall.
- Saurwein, K.-H., SPSS PC+: Eine anwendungsorientierte Einführung, 2. A. Bonn: Addison-Wesley.
- Schuboe, W., 1991: SPSS: Handbuch der Programmversionen 4.0 und 3.0. Stuttgart: G. Fischer.
- Steinhausen, D./Zörkendörfer, S., 1990: Statistische Datenanalyse mit dem Programmsystem SPSSx und SPSS-PC+, 2., völlig überarb. A. München/Wien: Oldenbourg.
- Wittenberg, RyCramer, H., 1992: Datenanalyse mit SPSS. Handbuch für computerunterstützte Datenanalyse Bd. 2. Stuttgart u.a.: G. Fischer (UTB).

Handbücher

- Norusis, M.J. & SPSS Inc., 1990: SPSS advanced statistics user's guide. Chicago: SPSS Inc.
- Norusis, M.J. & SPSS Inc., 1990: SPSS base system user's guide. Chicago: SPSS Inc.
- Norusis, M.J. & SPSS Inc., 1990: SPSS/PC+ advanced statistics 4.0 for the IBM PC/XT/AT and PS/2. Chicago: SPSS Inc.
- Norusis, M.J. & SPSS Inc., 1990: SPSS/PC+ 4.0 base manual 4.0 for the IBM PC/XT/AT and PS/2. Chicago: SPSS Inc.
- Norusis, M.J. & SPSS Inc., 1990: SPSS/PC+ 4.0 statistics 4.0 for the IBM PC/XT/AT and PS/2. Chicago: SPSS Inc.

- Schubö, W. u.a., 1991: SPSS. Handbuch der Programmversionen 4.0 und SPSS-X 3.0. Stuttgart: G. Fischer.
- SPSS Inc., 1990: SPSS Categories. Chicago: SPSS Inc.
- SPSS Inc., 1990: SPSS Data entry II for the IBM PC/XT/AT and PS/2. Chicago: SPSS Inc.
- SPSS Inc., 1990: SPSS/PC+ Graphics for the IBM PC/XT/AT and PS/2. Chicago: SPSS Inc.
- SPSS Inc., 1990: SPSS/PC+ Statistics 4.0 for the IBM PC/XT/AT and PS/2. Chicago: SPSS Inc.
- SPSS Inc., 1990: SPSS/PC+ Tables 4.0 for the IBM PC/XT/AT and PS/2. Chicago: SPSS Inc.
- SPSS Inc., 1990: SPSS/PC+ Trends for the IBM PC/XT/AT and PS/2. Chicago: SPSS Inc.
- SPSS Inc., 1990: SPSS/PC+ Map 4.0 from Mapinfo. Chicago: SPSS Inc.
- SPSS Inc., 1990: SPSS for the Macintosh: Operations guide. Chicago: SPSS Inc.
- SPSS Inc., 1990: SPSS for Unix: Operations guide. Chicago: SPSS Inc.
- Uehlinger, H.M./Hermann, D., 1992: SPSS/PC + Benutzerhandbuch. Band 1: Dateneingabe, Datenmanagement, Datenverwaltung und einfache statistische Verfahren, 2., akt. A. Stuttgart G. Fischer.

SPSS für Windows

- Brosius, G./Brosius, F., 1994: SPSS 6.0 für WindowsBaseSystem und Professional Statistics. Deutsche Ausgabe. New York: McGraw-Hill.
- Bühl, A./Zöfel, P., 1994: SPSS für Windows. Bonn: Addison-Wesley.
- Fieger, Toutenburg, H., 1994: SPSS Tables für Windows. Arbeitsbuch für Praktiker. München n.a.: Prentice Hall.
- Fieger, A./Toutenburg, H., 1995: SPSS Trends für Windows. Arbeitsbuch für Praktiker. München u.a.: Prentice Hall.
- Fieger, A./Toutenburg, H., 1995: SPSS Advanced Statistics für Windows. Arbeitsbuch für Praktiker. München u.a.: Prentice Hall.
- Fieger, A./Toutenburg, H., 1995: SPSS Professional Statistics für Windows. Arbeitsbuch für Praktiker. München u.a.: Prentice Hall.
- Foster, J., 1993: Starting SPSS/PC+ and SPSS for Windows, 2nd ed. New York: Sigma Press.
- Janssen, JTLaatz, W., 1994: Statistische Datenanalyse mit SPSS für Windows. Eine anwendungsorientierte Einführung in das Basissystem. Berlin u.a.: Springer.
- Kähler, W.-M., 1993: SPSS für Windows. Braunschweig: Vieweg.
- Lehnen, U., 1994: Datenanalysesystem SPSS für Windows. München: Oldenbourg.
- Norusis, M.J., 1993: SPSS for Windows, Release 6.0. Base System User's Guide. Chicago: SPSS Inc.

- Norusis, M.J., 1993: SPSS for Windows, Release 6.0. Advanced Statistics. Chicago: SPSS Inc.
- Norusis, M.J., 1993: SPSS for Windows, Release 6.0. Professional Statistics. Chicago: SPSS Inc.
- Schäfer, B./Jaschke, R., 1994: SPSS für Windows. Probleme, Beispiele, Lösungen. Braunschweig: Thomson.
- Schubö, W. u.a., 1994: SPSS für Windows. Stuttgart G. Fischer.
- SPSS Inc., 1993: SPSS for Windows, Release 6.0. Base System User's Guide. Chicago: SPSS Inc.
- SPSS Inc., 1993: SPSS for Windows, Release 6.0. Base System Syntax Reference Guide. Chicago: SPSS Inc.
- SPSS Inc., 1993: SPSS for Windows, Release 5.0 & 6.0. Tables. Chicago: SPSS Inc.
- SPSS Inc., 1993: SPSS for Windows, Release 6.0. Trends. Chicago: SPSS Inc.
- SPSS Inc., 1993: SPSS for Windows, Release 6.0. Advanced Statistics. Chicago: SPSS Inc.
- SPSS Inc., 1993: SPSS for Windows, Release 6.0. Professional Statistics. Chicago: SPSS Inc.
- Toutenberg, H.V.Fieger, A./Kastner, C, 1995: Statistik für Betriebs- und Volkswirte. Eine Einführung mit SPSS für Windows. München u.a.: Prentice Hall.

2.2 SAS (*»Statistical Analysis System«*)

- Aronson, M.V.Aronson, A., 1990: SAS (engl.). New York: McGraw-Hill.
- Aster, R./Seidman, R., 1991: Professional SAS programming secrets. Boston: Winderest.
- Bundschuh, W. u.a., 1991: Effektives Arbeiten mit dem Datenanalysesystem SAS. Mannheim: BI.
- Bundschuh, W. u.a., 1992: SAS. Mannheim: BI.
- Cody, R.P./Smith, J.K., 1987: Applied statistics and the SAS programming language, 2nd ed. Amsterdam: North-Holland.
- Dilorio, F., 1991: SAS applications programming. PWS-Kent.
- Dufner, J. u.a., 1992: Statistik mit SAS. Stuttgart Teubner.
- Götsche, T., 1990: Einführung in das SAS-System für den PC. Stuttgart G. Fischer.
- Herzberg, P.A., 1990: How SAS works, 2. A. Berlin u.a.: Springer.
- Kahler, W.-M./Schulte, W., 1993: SAS - Eine anwendungsorientierte Einführung, 3., neubearb. u. erw. A. Braunschweig: Vieweg.
- Küffner, H.V.Gogolok, J., 1994: Datenanalyse mit SAS. Handbuch für computerunterstützte Datenanalyse Bd. 4. Stuttgart u.a.: G. Fischer (UTB).
- Multrus, F./Bleicher, M., 1992: SAS. Eine praxisbezogene Einführung. Heidelberg: Hüthig.
- Nagl, W., 1992: Statistische Datenanalyse mit SAS. Frankfurt a.M.: Campus.

- Ortseifen, A., 1994: SAS. Thomson.
- SAS/STAT. Guide for personal computers. Version 6, 1987. Cary, NY: SAS Institute.
- SAS. Procédures guide for personal computers. Release 6.03, 1988. Cary, NY: SAS Institute.
- SAS. Language guide for personal computers. Release 6.03, 1988. Cary, NY: SAS Institute.
- SAS. User's guide: Basics. Version 5, 1985. Cary, NY: SAS Institute.
- Schuemmer, R./Gogolok, J./Ströhlein, G., 1990: Datenverarbeitung und statistische Auswertung mit SAS. Bd. 2: Komplexe statistische Analyseverfahren. Stuttgart u.a.: G. Fischer.
- SAS. User's guide: Statistics. Version 5, 1985. Cary, NY: SAS Institute.
- Schuemmer, R./Gogolok, J./Ströhlein, G., 1992: Datenverarbeitung und statistische Auswertung mit SAS. Bd. 1: Einführung in das Programmsystem, Datenmanagement Auswertung. Stuttgart u.a.: G. Fischer.
- Spector, P.E., 1993: SAS. Programming for researchers and social scientists. Newbury Park: Sage.
- Steinhausen, D./Zörkendörfer, S., 1992: Informationsbearbeitung und Datenanalyse mit dem Programmsystem SAS. München: Oldenbourg.

2.3 P-STAT (*»Princeton University Statistical«*)

- Küffher, H/Gogolok, J., 1994: Datenanalyse mit P-STAT. Handbuch für computerunterstützte Datenanalyse Bd. 3. Stuttgart u.a.: G. Fischer (UTB).
- Matiaske, W., 1990: Statistische Datenanalyse mit Mikrocomputern. Einführung in P-STAT und SPSS/PC. München/Wien: Oldenbourg.
- Voos, U.F./Gogolok, J./Küffner, H., in Vorb.: Einführung in P-STAT. 2 Bände. Stuttgart u.a.: G. Fischer.

2.4 BMDP (*»BioMedDical Package«*)

- Küffher, H./Wittenberg, R., 1985: Datenanalysesysteme für statistische Auswertungen. Eine Einführung in SPSS, BMDP und SAS. Stuttgart: G. Fischer.
- Pfeifer, A., 1988: Statistik-Auswertungen mit SPSS-X und BMDP. Ein Einstieg in die Benutzung der beiden Programmpakete. Stuttgart: G. Fischer.
- Snell, E.J., 1987: Applied statistics: A handbook of BMDP analysis. New York: Chapman & Hall.
- Wittenberg, R., 1993: Datenanalyse mit BMDP. Handbuch für computerunterstützte Datenanalyse Bd. 5. Stuttgart u.a.: G. Fischer (UTB).

- Bollinger, G./Hermann, A./Möntmann, V., 1983: BMDP Statistikprogramme für die Bio-, Human- und Sozialwissenschaften. Eine Beschreibung der Programmversionen 77-81. Stuttgart: G. Fischer.
- BMDP Statistical Software, Inc. (ed.), 1992: BMDP data manager manual. To accompany BMDP release 7. Los Angeles: BMDP.
- BMDP Statistical Software, Inc. (ed.), 1992: BMDP PC-92 user's guide. Los Angeles: BMDP.
- Dixon, W.J., (chief editor), 1992: BMDP **Statistical** software manual. Vol. 1 - Vol. 3. Berkeley: Univ. of California Press.
- Lüpsen, H., 1993: Einführung in das statistische Programmsystem BMDP in den Versionen 90 und 93 (7.0). Stuttgart u.a.: G. Fischer.

2.5 SYSTAT (»System for Statistics«)

- Blankenberger, S., 1994: SYSTAT für Windows. Stuttgart G. Fischer.
- Urban, D. u.a., 1992: Systematische Statistik für die computerunterstützte Datenanalyse. Ein Handbuch zum Programm-Paket SYSTAT. Stuttgart u.a.: G. Fischer.
- Schäfer, B., 1994: SYSTAT: Statistik unter Windows. Thomson.
- Wilkinson, L., 1990: SYSTAT: The system for statistics. Evanston, IL: SYSTAT Inc.

2.6 CSS: STATISTICA

- Engtler, U., 1993: Datenanalyse mit CSS: STATISTICA (Handbuch für computerunterstützte Datenanalyse Bd. 6). Stuttgart u.a.: G. Fischer (UTB)
- Nielsen, A./Loll, U., 1992: Einführung in das Statistik- und Grafikprogramm CSS. Stuttgart G. Fischer.

2.7 S-PLUS

- Chambers, J.M/Hastie, T.J. (eds.), 1992: Statistical models in S-Plus. Wadsworth.
- Süselbeck, B., 1993: S und S-PLUS. Eine Einführung in Programmierung und Anwendung. Stuttgart G. Fischer.

2.8 Sonstige Dokumentationen zu Spezialprogrammen

- Brockwell, P.J., 1993: Interactive time series modelling package for the SPARC-workstation. Berlin u.a.: Springer.
- Byrne, B.M., 1989: A primer of LISREL. Basic applications and programming for confirmatory factor analytic models. New York: Springer.

- Hennerkes, W.A., 1990: MAXDATA. A time series database system. Berlin u.a.: Springer.
- Jöreskog, K.G./Sörbom, D., 1993: LISREL 8: Structural modeling with the SIMPLIS command language. Hilldale: Erlbaum.
- Jöreskog, K.G./Sörbom, D., 1993: LISREL 8 & PRELIS 2 user's guide supplements. Hillsdale: Erlbaum.
- Liu, L.-M/Hudak, G.B., 1992: Forecasting and time series analysis using the SCA statistical system, Oak Brooks, Illinois: Scientific Computing Associates Corp.
- SPSS Inc., 1984: USERPROC LISREL: Using LISREL VI within SPSSx. Chicago: SPSS Inc.
- STATGRAPH, 1993. Heidelberg: tewi.
- STSC Inc., 1989: STATGRAPHICS user's guide version 4.0. Rockville: STSC Inc.
- Wishhart, R.D., 1993: CLUSTAN Version 3.2. Benutzerhandbuch. Stuttgart: G. Fischer.

2.9 Harvard Graphics

- Esser, W., 1991: Harvard Graphics 3.0 - Das Kompendium. Haar Markt & Technik.
- Esser, W., 1992: Harvard Graphics für Windows - Das Kompendium. Haan Markt & Technik.
- Huttel, K.P., 1992: Harvard Graphics für Windows 1.02. Bonn: Addison-Wesley.
- Huttel, K.P., 1993: Harvard Graphics 3.05: Grafik, Diagramme, Präsentationen. Bonn: Addison-Wesley.
- Schulte, G., 1992: Präsentationen mit Harvard Graphics für Windows. München: Francis.